

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 09:20:56 ; Search time 50.85 Seconds
(without alignments)
624.028 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774

Sequence: 1 MKSSFFKVFSTFAIFPLSM.....FELRGSSRYNYDLGAKYOF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36: *
1: /SIDS6/gcgdata/geneseq/geneseq/AA1986.DAT: *
2: /SIDS6/gcgdata/geneseq/geneseq/AA1981.DAT: *
3: /SIDS6/gcgdata/geneseq/geneseq/AA1983.DAT: *
4: /SIDS6/gcgdata/geneseq/geneseq/AA1984.DAT: *
5: /SIDS6/gcgdata/geneseq/geneseq/AA1985.DAT: *
6: /SIDS6/gcgdata/geneseq/geneseq/AA1986.DAT: *
7: /SIDS6/gcgdata/geneseq/geneseq/AA1987.DAT: *
8: /SIDS6/gcgdata/geneseq/geneseq/AA1988.DAT: *
9: /SIDS6/gcgdata/geneseq/geneseq/AA1989.DAT: *
10: /SIDS6/gcgdata/geneseq/geneseq/AA1990.DAT: *
11: /SIDS6/gcgdata/geneseq/geneseq/AA1991.DAT: *
12: /SIDS6/gcgdata/geneseq/geneseq/AA1992.DAT: *
13: /SIDS6/gcgdata/geneseq/geneseq/AA1993.DAT: *
14: /SIDS6/gcgdata/geneseq/geneseq/AA1994.DAT: *
15: /SIDS6/gcgdata/geneseq/geneseq/AA1995.DAT: *
16: /SIDS6/gcgdata/geneseq/geneseq/AA1996.DAT: *
17: /SIDS6/gcgdata/geneseq/geneseq/AA1997.DAT: *
18: /SIDS6/gcgdata/geneseq/geneseq/AA1998.DAT: *
19: /SIDS6/gcgdata/geneseq/geneseq/AA1999.DAT: *
20: /SIDS6/gcgdata/geneseq/geneseq/AA2000.DAT: *
21: /SIDS6/gcgdata/geneseq/geneseq/AA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4774	100.0	928	21	Chlamydia pneumoniae
2	4760	99.7	928	20	Chlamydia pneumoniae
3	2802.5	58.7	597	20	Chlamydia pneumoniae
4	2058	43.1	928	20	Chlamydia pneumoniae
5	2006	42.0	918	21	Chlamydia pneumoniae
6	2001	41.9	918	20	Chlamydia pneumoniae
7	1993	41.7	914	20	Chlamydia pneumoniae
8	1986	41.6	928	20	Chlamydia pneumoniae
9	1940	40.6	930	20	Chlamydia pneumoniae
10	1927	40.4	930	20	Chlamydia pneumoniae
11	1917.5	40.2	927	20	Chlamydia pneumoniae
12	1856	38.9	949	20	Chlamydia pneumoniae

13	1855	38.9	928	20	Chlamydia pneumoniae
14	1832	38.4	945	21	Chlamydia pneumoniae
15	1811	37.9	945	20	Chlamydia pneumoniae
16	1572.5	32.9	643	20	Chlamydia pneumoniae
17	1564	32.8	841	20	Chlamydia pneumoniae
18	1430.5	30.0	922	20	Chlamydia pneumoniae
19	1417.5	29.7	922	20	Chlamydia pneumoniae
20	1377.5	28.9	973	21	Chlamydia pneumoniae
21	1259	26.4	671	20	Chlamydia pneumoniae
22	1214	25.4	230	20	Chlamydia pneumoniae
23	1203.5	25.2	507	20	Chlamydia pneumoniae
24	1130	23.7	1132	20	Chlamydia pneumoniae
25	1128	23.6	1012	20	Chlamydia pneumoniae
26	1110.5	23.3	1013	20	Chlamydia pneumoniae
27	1098.5	23.0	1013	20	Chlamydia pneumoniae
28	1015.5	21.3	450	20	Chlamydia pneumoniae
29	918	19.2	610	20	Chlamydia pneumoniae
30	888	18.6	530	20	Chlamydia pneumoniae
31	834.5	17.5	483	20	Chlamydia pneumoniae
32	821	17.2	494	20	Chlamydia pneumoniae
33	792	16.6	427	20	Chlamydia pneumoniae
34	768	16.1	177	20	Chlamydia pneumoniae
35	708	14.8	294	20	Chlamydia pneumoniae
36	680.5	14.3	1617	20	Chlamydia pneumoniae
37	667	14.0	947	21	Chlamydia pneumoniae
38	667	14.0	1146	20	Chlamydia pneumoniae
39	646	13.5	969	20	Chlamydia pneumoniae
40	635.5	13.3	400	20	Chlamydia pneumoniae
41	635	13.3	279	20	Chlamydia pneumoniae
42	603.5	12.6	940	20	Chlamydia pneumoniae
43	556	11.6	1194	20	Chlamydia pneumoniae
44	551	11.5	924	20	Chlamydia pneumoniae
45	546	11.4	989	20	Chlamydia pneumoniae

ALIGNMENTS

RESULT	ID	1	ALIGNMENTS
Y94327	Y94327	Standard: Protein: 928 AA.	
XX	XX		
AC	Y94327		
XX	XX		
DT	11-AUG-2000	(first entry)	
XX	XX		
DE	Chlamydia pneumoniae 98kD putative outer membrane protein.		
KW	Chlamydia: antigen; vaccine; infection; outer membrane protein.		
XX	XX		
OS	Chlamydia pneumoniae.		
PN	W0200026237-A2.		
XX	XX		
PD	11-MAY-2000.		
XX	XX		
PF	29-OCT-1999:	99MO-GB03579.	
XX	XX		
PR	29-OCT-1998:	98US-0106070.	
PR	01-MAR-1999:	99US-0122066.	
PR	27-OCT-1999:	99US-0428122.	
XX	XX		
PA	(CONN-) CONNAUGHT LAB LTD.		
XX	XX		
PI	Muridin AD, Oomen RP, Dunn PL;		
XX	XX		
DR	WPI: 2000-365569/31.		
XX	XX		
XX	N-PSDB: A27021.		
PT	Novel Chlamydia 98 kDa putative outer membrane protein antigen, used		
XX	for vaccination and protection against Chlamydia infection		
PS	Claim 6; Fig 1; 93pp; English.		

XX The present sequence is the 98kDa putative outer membrane protein from
 CC Chlamydia pneumoniae. The genomic sequence was amplified using two
 CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome
 CC binding site, an initiation codon and a sequence close to the 5' end of
 CC the 98kDa putative outer membrane protein coding sequence. The 3' primer
 CC contains the sequence encoding the C-terminal sequence of the putative
 CC outer membrane protein and a BamHI restriction site. The stop codon was
 CC excluded and an additional nucleotide was inserted to obtain an in-frame
 CC C-terminal fusion with the Histidine tag. The PCR product was cloned
 CC into a eukaryotic expression vector (pCA-Myc-His) by restricting both
 CC the vector and the PCR product with NotI and BamHI and performing a
 CC ligation reaction. This expression vector was injected intramuscularly
 CC and intranasally into mice, which were subsequently inoculated with
 CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
 CC were lower than those of the controls. Thus the 98kDa putative outer
 CC membrane protein can be used as a vaccine to provide protection against
 CC Chlamydia infections, especially Chlamydia pneumoniae infections.
 CC The present polypeptide may also be administered orally to treat
 CC Chlamydia infection.

SO Sequence 928 AA:

Query Match 100.0%; Score 4774; DB 21; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKSFPKVFSTFAIFPLSMATETVLDSSAFEDNKNNGFVSREQDAGTYLEKGVN 60
 Db 1 mksfpkfvfstfaifplsmatetvldssafednknngfvsreqdagtylfxgvn 60
 OY 61 TLENIPGTGTAITSCENNTKGDLTFTGNGSLFQTVADGTVAGAVNSVVDKSTFEI 120
 Db 61 tlenipgtaigtatkcfnntkgydlftcngnslftqvadgvtvagaavnsvvdksctfi 120
 OY 121 GFSSLSFASPGSSITTGKGAVSCSTGSLSTKNVSLFSKNEFTDNGAITAKTSLSLG 180
 Db 121 gfsslsfaspssitlgkgaavscstgslstknvslfsknfctdngaitaktslsltg 180
 OY 181 TTMSALFSENTSSKKGAIQTSALITTTGNOGEVFSFSDNTSSDGAALFEASTISNNA 240
 Db 181 ttmsalfseentsskkgaiqttsalittgnogevfsfndntssdgaalfeastisnna 240
 OY 241 KVSFTDNKVTGASSTTGDMGCAICAYKSTDPKVTLTGNOMLFSNNTETGAIIV 300
 Db 241 kvsftdnkvtagsssttgdmgcaicaykstdpkvtltgnoqlfsmntetggaiv 300
 OY 301 KKLKIASGGLTFSRNSVNGGTAPKGAIAIEDSGELSLADSGDIYFLGNTVTSTPGT 360
 Db 301 kklkiasggltfstrnsvnggtapkgaiiaiedsgelsladsdgiyflgntvtstpgt 360
 OY 361 NRSSIDIGTSAKMTALRSNAGRAIFYPPIITGSSITVDLKNVETPADALQYTGMI 420
 Db 361 nrssidigtakmtalrsnagraifyppitgssitvdlknevtpadalsqlytgmi 420
 OY 421 FTGKELSTFEADSKNLTSLLOPYTLISGGLSLKHGVTLLQOAFYQOADSRLKEDYGT 480
 Db 421 ftgkelsftheadsknltsllopytlisggltslkhgvtllqoafyqoadsrlkedygt 480
 OY 481 LEPADISTNNLVINISSIDGAKKAIETKATSKNLTLSGTTLLDPTGYENHSLRNP 540
 Db 481 lepadistnnlviniissidgakkaiekatsknltlsgttllldptgyenhslnrp 540
 OY 541 QSYDILKASGTVSTAVTPPPINGEKHYGOGTWGPYVNGIGTASTTATNMKTGYI 600
 Db 541 qsydilkasgvtstavtpppingekhygogtwgpyvngigtasttatnmktyi 600
 OY 601 PNPERSGVPSNINWAFIDISLHYMETANEGLOGDAPFACAGISNFFHDSKTRRG 660
 Db 601 pnpersgvpsnlnwafidislyhmetaneglogdafacagisnffhdscktrrg 660
 OY 661 FRHLSGGYVIGNLTHTCSDKILSAFQLEGRDRDYFVAKNOGYVGGTLTYOHNETYIS 720

Db 661 frhlsggvlggnlhtcsdkilsaafqlegrdrdyfvakngtyvggtylyqhnetyis 720
 OY 721 LPCLRPCCSLSYPTPELPVLFSCNLSYTHDNLKTKYTPPVKSSMKNDSFALFEGGR 780
 Db 721 lpclrpccslsyptpelepvlfsnlsythdnlktrktyppvkssmknndsfalfeggr 780
 OY 781 APICLDSALFEQYPMKIQFYVAHOGFKEQTEAREFGSSRLVNLALPIGIRFDKES 840
 Db 781 apicldsalfeqypmkilqfyvahogfkegtearefgssrlvnlalpigirfdkes 840
 OY 841 DCQDATTNLTGYVDLVRSNPDCTTLRLSGDSMKFCTGNLRLQALVIRAGNFCFNSN 900
 Db 841 dcqdattnltgyvdlvrsnpdcttlrlsgdskmfctgnlrlqalviragncfnsn 900
 OY 901 FEARQSFEEIRGSSRNPNVDLGAKEYF 928
 Db 901 fearqsfeelrgssrnpnvdlgakgyf 928

RESULT 2

W88421
 ID W88421 standard; Protein; 928 AA.

AC W88421;

DI 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp8.

KW Omp8; outer membrane protein 8; surface exposed protein; antigen;
 infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.
 (CHRI/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 Mygdal P;

XX WPI, 1999-105610/09.

DR N-PSDB; X06820.

PT Species-specific test for identifying mammals infected with
 Chlamydia pneumoniae - comprises detecting antibodies specific for
 outer membrane proteins of C. pneumoniae or nucleic acids encoding
 these proteins

XX Claim 7; Page 53-55; 115pp; English.

XX This polypeptide comprises the novel 90.0 kDa surface exposed
 CC protein Omp8 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06820) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins; Omp4-Omp5
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

XX Sequence 928 AA;

Query Match 99.7%; Score 4760; DB 20; Length 928;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 925; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 1 MKSSPKKFEVFETFAFPMTATETVLDSASFDGNKNGNSVRESQDAGTTLVFKGNV 60
DB 1 mksspkkfvtfatfaplsmlatevldssafdgknknfnsvresqdaqctylfkgnv 60
OY 61 TLENIPGTGTAITKSCFNNTKGLFTFTGNGNSLFPQTVAGTACAAVNSVVDKSTTFI 120
DB 61 tlenipgtgtaitekcfntksgdlftfgngnsllftqvadgtvagaavnsavdksttfi 120
OY 121 GFSSLSFIASFGSSITTKGAVSCSTGSLTKNNSLFSKNFTSDNGAITAKTSLTG 180
DB 121 gfsslsfiasspgssiltkgavscstgsltknnsllfsknftcdngaitaktsltg 180
OY 181 TTMSALFSENTSSKKGAIOTSDALTITNGOEVFSDNTSSDGALEFTEASVTISNNA 240
DB 181 ttmsalfesntsskkgaiotsdaltitngoevfsdntssdgaalefteasvtisnna 240
OY 241 KVSFIDNKVTCASSSTTGDMSGAICAYKSTDTKVLITGNOMLFSNNTSTTAGAIYV 300
DB 241 kvsfidnkvtcasssttgdmsgaicaykstdtkvlitgnomlfsnntsttagaiyv 300
OY 301 KKLEIASGGLTFRSNSVNGTAPKGAIAIEDSGELSDSGDIVFLGNVTSTPFGT 360
DB 301 kkleiasggltfrrnsvngtapkgaiaiedsgelsdsgdivflgnvtstpfgt 360
OY 361 NRSIDLTSAKMTALRSAGRAIFYDPITTGSSSTVTDVLKVNERTPADSALOYTGNI 420
DB 361 nrsidltsakmtalrsagraifydpitgssstvtavlknvrtpadsaloytgnii 420
OY 421 FTGEKLESTEAADSKNLTSLQPTLSGGLSLKHGYTLQTOAFTQOADSRLNDVGT 480
DB 421 ftgeklesteaadsknltslqptlsggslslkhgytlqtoaftqoadsrlndvgt 480
OY 481 LEPADTSTNNLVINISSIDGAKKAKIEFKATSKNLTSGTTLDPGTGYEENSLNPN 540
DB 481 lepadtstnnlvinnissidgakkakiefkatsknltsgttdlpgtgyeenlnp 540
OY 541 QSYDILELKASGTVTSTAVTPDPIMGEKFXHYGQGTWGPVWGTGASTTATFNMKTGY 600
DB 541 qsydilelkasgvtvstavtpdpimgkfxhyggtwgpvwgtgasttatfnmktgy 600
OY 601 PNPERIGSLVNSLMAIDISSLHYLMEETANEGLOGRARWACGLSEFFKDKSTKTRG 660
DB 601 pnperigslvnslnmaidisslhylmetaneglograrwacglseffkdkstkrtrg 660
OY 661 FRHLSGCVIGVIGNLTHTCSDKILISAACOLFGDRDRDYFAKNQGYVGGTLYVONHET 720
DB 661 frhlskgvlgvignlthtcsdkilisaacolfgdrdrdyfaknqgyvggtlyvohet 720
OY 721 LPKCULRPSLSYVPTETPLVPSGNLSYTHNDLTKYTTYPVYKSGNDSFALFEFCGR 780
DB 721 lpculrpslsyvptetplvpsgnlsythndltktytpvyksgndsfalfefcgr 780
OY 781 APICLIDSEALFEQYWPMPKLOFYVAHOGCFKEOGTEAREPSSRPVNALITGRPKES 840
DB 781 apiclidealseqywpmpklifyvahogcfkeogtearepssrpvnalitgrpkess 840
OY 841 DCODATYNTLGYTVDLVRSNPDCTTTLRISGDSKWTFGTINLAROALVLARGNHCFES 900
DB 841 dcoatynltlgytvdlvrsnpdcttllrisgdskwtfgtinlaroalvlargnhcfes 900
OY 901 FEAFSQFSELRGSSRNINVDLGAKYOF 928
DB 901 feafsqfselrsgssrninvdlgakyo 928

```

DB 901 feafsqfselrsgssrninvdlgakyo 928

RESULT 3
ID Y34611 standard; Protein: 597 AA.

XX AC Y34611;
XX DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

PN 03-JUN-1999.

PE 20-NOV-1998; 98WC-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GSET) GENSET.

PI Giffels R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 632-633; Disclosure; 1912pp; English.

CC Y34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.

SO Sequence 597 AA;

Query Match 58.7%; Score 2802.5; DB 20; Length 597;
Best Local Similarity 93.8%; Pred. No. 6,5e-178;
Matches 560; Conservative 5; Mismatches 21; Indels 11; Gaps 3;

```

OY 149 LSLTKNNSLFSKNFTSDNGAITAKTSLTGTTMSALFSENNSSKKGAIOTSDALTIT 208
DB 1 lsltknnsllfsknftsdngaitaktsltgttmsalfseennsskkgaiotsdaltit 208
OY 209 GNOGEVFSFSDNTSSDGAALFTEASVTISNNAKVSFIDNKVTCASSSTTGDMSGAICAY 268
DB 61 gnggevfsfntssdgaalfteasvtisnnakvsvfndknvtcasssttgdmsgaicay 268
OY 269 KTSTDITVTLTGNOMLFSNNTSTTAGAIYVKKLEIASGGLTFRSNSVNGTAPKGA 328
DB 121 ktstditvltlgnomlfsnntsttagaiyvkkleiasggltfrrnsvngtapkga 328
OY 329 IATDSELSLSDSGDIVFLGNVTSTPGTNRSSIDLTSAKMTALRSAGRAIFYD 388
DB 181 iatdselslsdsgdivflgnvtstpgtnrssidltgsakmtalrsagraifyd 388

```

QY	389	PLTSSSTVTMDVLKVNETPNAPSALQYTGNIIFGKISLSEADSKNLTKLQPTTIS	448
XX			
Db	241	plttssstvtvdlkvnetpnapsalqytgnliffgekiseteaadsknltskllqpttIs	300
QY	449	GCTLSLKHGVTLLQOTAFATQOADSRLKEMOVGTTLEPADSTNNLVINISSIDGAKKAKIE	508
XX			
Db	301	ggctlslkhgvtllqtaftqgadsrlkemdvgttlepadstnnlvinnissidgakkakie	360
QY	509	TRATSKNLTSLSTITLLDPTGTFEENHSLRNPQSDIIELKASGTVTSTAVTPDPIWGEK	568
XX			
Db	361	tratsknltlssttllcllpdptgfeyhbslrmpgsydllelkasgtvstavtpdplmgek	420
QY	569	FHYGOGTGWGPVWOTGASTATFEMWTKGTGYPPIPERIGSLVPSLNNAPRIDSLSHYLM	628
XX			
Db	421	fhygogtgcwplwgtgastatfemwtktylppperigslvpslwnatridslshylm	480
QY	629	ETANEGLOGDRAFWACAGLSNFFHKDSTKTRRGFPHLSGGVYIGGNLTCSDKILISAFCQ	688
XX			
Db	481	etaneaglgdratfwacagslnffhkdstktrrgfphlsggvyiggnltcsdkilisaafcg	540
QY	689	LEGRDRDVFVAKNKGSTYGGGTLVYQNHMYRIS-----LPCKLR-PCSLSTYP	734
XX			
Db	541	lgrdrdyfvaknkgstlrnslpdrnlylsslqtlatlvclsydsclfrxP	597
RESULT	4		
ID	W88418	W88418 standard; Protein; 928 AA.	
XX		W88418:	
XX		26-APR-1999 (first entry)	
DT		Chlamydia pneumoniae surface exposed protein Omp5.	
XX			
DE		Omp5: outer membrane protein 5; surface exposed protein; antigen;	
XX		infection; diagnosis; vaccine; atherosclerosis; asthma.	
KW		Chlamydia pneumoniae.	
OS			
XX		WO858953-A2.	
PN			
XX		30-DEC-1998.	
PD			
XX		19-JUN-1998; 98WO-DK00266.	
PF			
XX		23-JUN-1997; 97DK-0000744.	
PR			
XX		(BIRK/) BIRKELJUND S.	
PA		(CHR/) CHRISTIANSEN G.	
XX			
PI		BirkeLund S, Christiansen G, Knudsen K, Madsen A;	
XX		Myind P;	
DR		WPI: 1999-105610/09.	
XX		N-PSDB; X06817.	
PT		Species-specific test for identifying mammals infected with	
XX		Chlamydia pneumoniae - comprises detecting antibodies specific for	
CC		outer membrane proteins of C. pneumoniae or nucleic acids encoding	
CC		these proteins	
CC		Claim 7; Page 43-45; 115pp; English.	
CC		This polypeptide comprises the novel 97.2 kDa surface exposed	
CC		protein Omp5 of the human respiratory pathogen Chlamydia	
CC		pneumoniae. Its amino acid sequence was deduced from DNA (see	
CC		X06817) isolated from a C. pneumoniae expression library. The	
CC		invention provides 12 novel surface exposed proteins, Omp4-Omp15	
CC		(see W88417-28), and nucleic acid sequences encoding them (see	
CC		X06818-27). A new species specific test is claimed that is used	
CC		to identify mammals (including humans) infected with Chlamydia	
CC		pneumoniae. The test comprises detecting antibodies specific for	

CC	Query Match	Best Local Similarity	Matches 444;	Conservative 142;	Mismatches 320;	Indels 44;	Gaps 20;
CC	1 MKSSPFKEVF-STFAIF--PLSMATETVLDDSSAFDKNK-GNFVSRESQEDAGTYVL 55	43.1%;	Score 2058;	DB 2058;	Length 928;		
CC	1 MKSgfwLWlslstlaactscstvfataenlpspsdfgstntgcytknt--ttgldt 58	46.7%;	Pred. No. 2.8e-126;				
CC	56 FKGNTVLENIPEGTGAITSCFNNTKGDITFFGNGSLFQTVDAGYAAGAVNSVYDK 115						
CC	59 lqgdltlqnl-dsaaltkyctdctdeslfsqkykslfnlks-saeagal-vstldk 115						
CC	116 STTFIFSSLSLFIASPGSSITF--CKGAVSCGSLSLTKNLSLKFNFSDNGATYA 173						
CC	116 nslsgfssltllaapsvltlpskqgavkg-qgltldmgllflfkqdyceengaalst 174						
CC	174 KRLSLGTGTMSLFSEBNTSS--KKGAIGTISDALITGNGOEVFSNNTSSDGAIAFT 230						
CC	175 knlslkngstsisfegnkssatgkkggalcagvtldtnaptlfnsliaaagaans 234						
CC	231 EASVTSNNAKRSFIDNKVTGASSTGTGMSGCAICAKTSDTAVTLTGOMLFNNT 290						
CC	235 tgncltgnltslviefenst----atag--nggal-----sgdaavltslgnsvfsgnq 283						
CC	291 STTAGAGIYVKKLELAS--GGITLFSRNSVNGTAPKGAIAIEDSGELSLSDSGDIVF 348						
CC	284 avangagalyakklltaasggggisfsmlygltagngaisllaagecslaaegdltf 343						
CC	349 LGNTVSTTP--GTNNSIDLGSAKMTLRASAGAIIFYEYPTITGSSITYTDLKVNET 407						
CC	344 ngnaavatlcpqtknsndlgctaklthlraagnsllfypilntaadacltlnlka 403						
CC	408 PADSALOYTGNIIFGKELSETEADSNKLSKLLQPTLGGTSLKHGVTLOTQATQ 467						
CC	404 dagnstdygsivfageklsedeakvadalntslkqptlagnlrlkrytldtkcgt 463						
CC	466 QADSLLEMDVGTLEPA--DSTINLVINISSIDGAKAKIETKATSKNLLSGTITLLD 526						
CC	464 tagssvlnadagltlkasteevltlglslpvsllgeqkvvlaaasaaknvaslpilld 523						
CC	527 PTGYTEVHSHSLNPQSYDILLELkAGVYTAIVPDPIMGEKFNHGYGTGMPITW--- 582						
CC	524 ngnaayenhdlktgdfsfvqalsagltcttdpavrcvaprthnygygtvg--ntwddt 582						
CC	583 -GTGASTATATENWTKTGIIPNPERIGSLVPSNLNAFIDISLHMETANEGLQGDRAF 641						
CC	583 astpklrtlatlawnltgylpnperegplvpmsllwgsfedaigaivlteralsatlcsdrgf 642						
CC	642 WCAGLSNFFEHSDSTRKTRGPFHLGSGYVIGGNLHHCPSKILISAAPCOLFGGDRDYPAKN 701						
CC	643 waagvaanlldkdkygekkkynhksagysyigaagatceenlslsfqqlfgsddkflvaakn 702						
CC	702 OGTVYGGTLVYQHNETVYISLPCRLPSCLSVYP--TEIPVLFSGNLSYTHTDNLTKY 758						
CC	703 htdtyagafgyqh---ltescgfigclldklpsswnshkplvleagqlaysvndlktky 758						
CC	759 TTYPIYKSSMGNDSSALEFGGRAPICLDESKALFEQYMPFMKLOVYIAHOESFKOGEAR 818						
CC	759 layevkkswnnnaftmmllqasshyapeylhncftdyapyiklnltlylrqdsfisekgtegr 818						
CC	819 EFGSSRLVNALPLGIRFDKESCDODAYNNLTLYVDLVNSNDPCTTTRISDSMKTFF 878						


```
Db      819 sfdsnlnflslpkyvfkfscndfslsyalltlyvpdlrlnbpccktaivlsqaswey 878
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      879 GTNLARQALVLRAGNHCFNSNEAFSQFSEFLRGSSRNYNVDLGAKYQF 928
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      879 annlarqalvragshyafspmfefvlgqfvefygssrlynvdlgkfqf 928

RESULT  5
ID      Y69369
        Y69369 standard; Protein: 918 AA.
XX
AC      Y69369;
XX
DT      19-JUN-2000 (first entry)
XX
DE      Amino acid sequence of the CPN100395 polypeptide.
XX
KM      CPN100395; Chlamydia infection: immune response; vaccine.
XX
OS      Chlamydia pneumoniae.
XX
PN      WO200011183-A2.
XX
PD      02-MAR-2000.
XX
PF      18-AUG-1999: 99MO-IB01449.
XX
PR      20-AUG-1998: 98US-0097187.
XX
PR      20-AUG-1998: 98US-0097188.
XX
PR      20-AUG-1998: 98US-0097189.
XX
PR      20-AUG-1998: 98US-0097190.
XX
PR      20-AUG-1998: 98US-0097195.
XX
PR      20-AUG-1998: 98US-0097196.
XX
PR      20-AUG-1998: 98US-0097197.
XX
PR      20-AUG-1998: 98US-0097191.
XX
PR      17-AUG-1999: 99US-0376770.
XX
PA      (CONN-) CONNAUGHT LAB LTD.
XX
PI      Mordin AD, Oomen RP;
XX
DR      MPI: 2000-224703/19.
XX
DR      N-PSDB: 261509.
XX
PT      Novel antigens and corresponding DNA molecules that can be used to
XX
PT      prevent, treat and diagnose disease caused by Chlamydia infection in
XX
PT      mammals, especially humans -
XX
PS      Claim 19; Fig 15-E; 201p; English.
XX
CC      Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
XX
CC      are present in the bacterial membrane structure, in the external
XX
CC      vicinity of the membrane structure. In the inclusion membrane
XX
CC      structure, in the external vicinity of the inclusion membrane structure,
XX
CC      and in the cytoplasm of the infected cell. The polypeptides may be
XX
CC      used to prevent, treat and detect the presence of Chlamydia infection
XX
CC      and/or the presence of Chlamydia in a sample. The polypeptides may
XX
CC      also be used to induce an immune response in a mammal. The vaccine
XX
CC      vector comprising the polynucleotides is used to induce an immune
XX
CC      response in a mammal. Antibodies directed against the polypeptides
XX
CC      may also be used therapeutically to treat and/or prevent a Chlamydia
XX
CC      infection.
XX
SQ      Sequence 918 AA:

Query Match      42.0%; Score 2006; DB 21; Length 918;
Best Local Similarity 45.6%; Pred. No. 7.7e-125;
Matches 432; Conservative 150; Mismatches 316; Indels 50; Gaps 20;

OY      1 MKSSFKFVSTFAIPLSMT---ATEYLDSSASFDGN-KGNNSVRSQEDA-GTTY 54
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db      1 mssffllllssslapllmsvsadaatltsrdsyngdtctftfpaaksdsagty 60
OY      55 LFKGNVTLLENIPOTGTAIRKSCFNNTKGDLTFTGNGNSLLFQTVDAITVAGAANSSVVD 114
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 lldgdrvsiaq-egkqslstscfsmagnlftfngfshfchfnllsslvagvvnstlaas 119
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      115 KSTFTGFSSLSLFIAPSGSSITTTGKAVSCSTGSLSTLRNVSILFSKNSTDNCAITRK 174
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120 gltkisgfsllmlaapr---tlgkgaikltg-lyfeisignldnennssenggalnck 175
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      175 TSLTGTMTSALFSENTSSKKGAIOISDALRTTGNQGVESFDNTSSPGAIFFEAIV 234
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      176 tslstgstrfvalignssqggaiysagdsvisenagllsfmnsatssgaiasegnl 235
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      235 TISNNAKVSFIDNKVTGASSSTTGDNSSGALCAVY--TSDTRKVTLTGNQMLFSNNTST 292
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      236 vlsnngnlffdgkat-----tngaidcnkagandpdlitlsneslhflnntag 286
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      293 TNGAIIYVKLELASG-GLTLFSRNVNGCTAPKGAIAIEDSGELISADSGDIYFLGN 351
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      287 nsqgaalytkkllvssgrgylfsmkaanat-pkqgaiaildsgelsadignllfegn 345
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      352 TVTSTT--PGTNRSSIDLTGSAKMTALRSAAGRAIFYDPITTGSTTVTDVLYKNETP 408
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      346 t-stltgspasvvrnaidlasnakflnlatrgnkylfydplt---ssgatklslnked 401
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      409 ADSALOYTGNITFTGKLESTFEAADSKNLTSKLQDPVTLSGGLSLKHVVTQTOAFIQO 468
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      402 agsgnlyegylyfsgeliseekkpndlkstftgvevlaagalykddvrvvanltbv 461
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      469 ADSRLMDVQTTLE-PADSTINNLVINISSIDGAKAKIEIKATSKNLTLSGTTILLDP 527
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      462 egskvwmvgdgttfesaeevtnlgnlndisdgcnkalikaaskaakdvalsgplmiva 521
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      528 TGTTFENHSLRNPQSDIILEKASCTVSTAVTPPDIMGEKHYGQGWGPVW---GT 584
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      522 qgnlyehnlsgqvfpfllsagcmtdtdpdlpdlcthhygygnwn-lywvddat 580
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      585 GASTTAFNMWTKGYIPNERIGSLVPSNLSMNAFIDISSLHLMETNANGLOGDRAFWCA 644
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      581 aktncaltwtkykpnrpergplvpnslwgsfydvrsigslnmdstsslsstcnlws 640
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      645 GLSNPFHKDSTKRGFRRLSGGVYVGNLHTCSDKILSAAFCOLFGRPDRDYFVAKNCGT 704
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      641 gladflhedqkngqrsyrssagyalggfftasenfftfacqlfygkxhllvaknhh 700
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      705 VYGGTILYQH---NETYISLPCKLRPCSLSYPTLEIPVLFSGNLSYTHTDNDLKITYTY 761
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      701 vygamsyrhlgesktlakl-----lsgnsdsipfinafayghdnmtktytg 752
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      762 PTVKSGMNDSEFLPEFGGRAPICLD-ESALFEQVPMFKLOEVVAHOEGFKEGTEAREF 820
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      753 spvksgwgdafgieeggaipvaagrrswdthpdlfnlmlayahqndfknegrsf 812
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      821 GSSRLVNLALPIGIREDKESDCODATYNTLTGYTVDLVRSNPDCTTTLRISGDSWKTEGT 880
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      813 gseedlnlavpyikfektfd--kstydisiaypvdvtrndpgcttllmvsdswstcgt 870
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY      881 NLARQALVLRAGNHCFNSNEAFSQFSEFLRGSSRNYNVDLGAKYQF 928
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      871 slsrqallvraqnhafnafevfesqflevelrgssrsyaldlgigtgi 918
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT  6
W88422
ID      W88422 standard; Protein: 918 AA.
XX
AC      W88422;
XX
DT      26-APR-1999 (first entry)
XX
DE      Chlamydia pneumoniae surface exposed protein omp9.
XX
```


PI Mybind P:
 XX WPI: 1999-105610/09.
 DR N-PSDB; X06828.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Disclosure: Page 78-80; 115pp: English.
 XX
 CC This polypeptide is described as a subsequence of the claimed
 CC novel surface exposed protein Omp5 (see W88418) of Chlamydia
 CC pneumoniae, a human respiratory pathogen. The invention provides
 CC a new species specific test for identifying mammals (including
 CC humans) infected with C. pneumoniae. The test comprises detecting
 CC antibodies specific for surface exposed proteins Omp4-Omp15 (see
 CC W88417-28) or detecting nucleic acid fragments encoding them (see
 CC X06816-27), especially by PCR. The proteins are also used in the
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
 CC and proteins can also be used in the immunization of mammals, the
 CC nucleic acids being particularly useful as DNA vaccines for
 CC effecting in vivo expression of antigens. The vaccines may also
 CC prevent atherosclerosis and bronchial asthma, which are possibly
 CC associated with C. pneumoniae.
 CC
 XX
 XX Sequence 914 AA:
 SQ
 Query Match 41.7%; Score 1993; DB 20; Length 914;
 Best Local Similarity 46.2%; Pred. No. 5,5e-124;
 Matches 432; Conservative 141; Mismatches 319; Indels 44; Gaps 20;
 QY 1 MKSSFPKVF-STFAIF--PLSMIAETVLDSASFDGKN-KNFSVRESQEDAGTTL 55
 DB 1 mksqfswlsvtlactscetvfaaenigpsdfsgntgtycpknt--ctgidyt 58
 QY 56 FKGNVTLENIPTGTATTKSCFNNTKGDITFTNGNSLLPQTVAGVVAAGAAVSSVVDK 115
 DB 59 lqgdlitqln-gdssaalqkqfscdleslfaagkyisflnks-aagaaal-svtdtk 115
 QY 116 STTFIGFSSLSFIASPCSSITF--GKAVSCSTGSLSLTKNVSLFSGNFSTDNGAITA 173
 DB 116 nlsitgfsstflaapsvltlpsgkavkcg-gdlfdmngllfkqdyeeengagalt 174
 QY 174 KTLUSLTGYMSALPSENTSS--KKGAIOTSDALTTGNOGEVFSFSDNTSSDGAIFT 230
 DB 175 knlskntsgsisfegnkksatgkkgacatgvtldntlnaplfanlaaagaalns 234
 QY 231 EASVTISNNKAVSFIDNKVYGASSSTTGDMGGAICAKTSTDKKVLTTGNOMLFEFNNT 290
 DB 235 tgnctlgnsisvtsensvtc---atag--ngaal-----sgdadvtisngsvstisng 283
 QY 291 STTAGAIIYKKLELAS--GGLTFSRNSVNGTAPKGAIAIEDSGELSLASDGDIVF 348
 DB 284 avanggaiaiyakklllaagsgggagisfnaiygttagngagtsllaagcstsaagdltf 343
 QY 349 LGNVITSTP-GTNRSSIDGTSKMTALSAAGRAIYFDPRTGSGSTVTVDLKNET 407
 DB 344 ngnaivaltqctkrnsidigtstaklnlraishsifdyplantaadastldlnlka 403
 QY 408 PADSALOYTGNIPTGEKLESTEADSKNLTKLOPVTSGGTSLKHGVTLOTQAF 467
 DB 404 dagnstlsgsivtsgeklseadeakvadnlstlkipvltlagnlvykryvltldtkftg 463
 QY 468 QADSRLEMDVGTLEPA-DTSTINNLVINISSIDGAKKAKIETAKSKNLTLSGTFITLD 526
 DB 464 tagsvimidagttlkasteetrlglslpvdslqegkvvlaaasaaknvalspilild 523
 QY 527 PTGTFYNNHSLRNQSYDILELKRSGVTSTRAVTPDIPMGKRPYGGOGMGPVW---- 582
 DB 524 ngnaayenhdlgkqdfsfvqlslagcattcdvavptvatphtlygygwtgw-mtwvddt 582

QY 583 -GTGASTTATFNWTKTGYPIDNPERISSLYPNLSMNAFIDISSLHYMETANEGLQGDRAF 641
 DB 583 aslprkktlatwntngylnpqrqplvpnsiwsfsdigaivlersaltilcddrgf 642
 QY 642 WCGLSNFFFKDSTKRRCGRHLSGGVIGGNLHTGSDKLTLSAFOQLFGRDDIYVAKN 701
 DB 643 waagvanfildkdkkgekrkyrhksygyaagaaqtsenlslafqqlfsgdxfivakn 702
 QY 702 QGTIVYGGTLYOHNETYISLPCRLRPSLSYVP---TEIPVLFGNSLYHTNTDLTKTY 758
 DB 703 hdttyagafiyqh-----lccsgfifglclldklpgswshkplvlegqdayshvndlky 758
 QY 759 TTYPTVYKSGWSDSFALFEGGRAPICLDESALREOYMPENKLOFYVAHDEGFEQGTAR 818
 DB 759 tayevekswgnafnmmlgaashsypeylhcdityapyikllylryqdsfsekgtegr 818
 QY 819 EFGSSRLVNLALPGRIFRDESCODATYVLTGYVNDVRSNPDCTTTLRISGDSWKTP 878
 DB 819 sfdsnlnfnslpkyvxfekfsdncdtsydltsyvpdlirnpkcttalvisgaewely 878
 QY 879 GTNLARQALVLRAGNHCFNSNEAFSQSFELRGS 914
 DB 879 annlarqalvragshyafspmevlgqfvevrgs 914
 RESULT 8
 W88423
 ID W88423 standard: Protein; 928 AA.
 XX
 AC W88423;
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed protein Omp10.
 XX
 KW Omp10: outer membrane protein 10; surface exposed protein;
 KW antigen: infection; diagnosis; vaccine; atherosclerosis; asthma.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN M09858953-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98MO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHR/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mybind P;
 DR WPI: 1999-105610/09.
 DR N-PSDB; X06822.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Claim 7: Page 60-62; 115pp: English.
 XX
 CC This polypeptide comprises the novel 98.4 kDa surface exposed
 CC protein Omp10 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X068422) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia

Dd	119	lItfgfmslfsiaepgtvtvasgktls-sagallhltngltlfsqvnseanmgga	177
Qy	174	KLtSLtGtTtMAlEsrEntSSrKKGaIoTSDAlTtTGNOGvEsFDntSDSGaItTEAs	233
Dd	178	KLtIsgrtstltfmsakklggaiysaaaslsngtltgvltnmhkgtgaggajlfses	227
Qy	224	VTISNAKAVSTIDNKVYGASSTTGDMSGAICAKYKSTDTKVTlTNGNMLFSNNTSTT	293
Dd	228	ssltqnsllftfsgtltataag-----kgaalyekkgteptltlsnksltlaensvvt	291
Qy	294	AGCAIYYVKtLLAGGtLTFrSNsvNGtPARKGaIKIEDSGELtsADSGdlVlFGNTV	353
Dd	292	qggatcahgdlsaaqpllfennrcgnttaagkggaiaaasgsilsaagddltfignlt	351
Qy	354	TSST- PGTNRSSIDLGTSAKMTALRSAGRAIYFDPI--TTGSSVTYVdLVKMETPA	409
Dd	352	tsesptetnaaiylgssakcltnrraaggqgiyllydplaantbas---dvltlqpds	407
Qy	410	DSALOYGNIIETGEEKISETFEPAASKNLTSLLOPVLtSGTSTSLIKNGVtLQOARTQOA	469
Dd	408	nspldysctvlytsektlsadeakaadftslkplxlaagclalxgveldvngtltqce	467
Qy	470	DSRLMDVGTTLLEBDSTTIN--NLVINISSIDAKKAKIETKATSKNLTlSGTITLDP	527
Dd	468	gstllmqgtytk-adeaaisltklvldlsalegnksvsetaganxtlttsplrvfids	526
Qy	528	TGTFYENSLRNPOSYDLLEKAS-----GTVSTAVTDPDPIKGEFHYGYOGTWG	578
Dd	527	stgnlyesltlnqatfpvlvfaataasdllydalllsvpdrpep----hygyqhw	580
Qy	579	PIVWGtGASTATFNTKTYGTPYPERIGSLVPSLlMNAFIDISLHYLMETANEGLQD	638
Dd	581	atwadtstaksgtmtwtvtygnpneratavpdsllwasftdrtltlqmtsganslyq	640
Qy	639	RAFNCAGLSNEFHDSRTKTRGRFHLBSGYVIGGNLTGCDKILSAFCOLFERDDPY	698
Dd	641	rglaasgaanfthkdksgltqaftkhsxygylvgsaaedfseentfsavcglfidxdlti	700
Qy	699	AKNGTAYGGtLlyVOHNETVtSLPCKRtPC--SLSYVPTPVLFSGNLSYtTDDDLKT	756
Dd	701	ventshoylaelyqhnaflgblp---mpstfgstlmdlkdlrlllnaqslsytlkmdnt	757
Qy	757	KYTYTPYVKGSGMDSRFLEREGGAPtCL-DESALEFGQVYPMFKLQVtYAHQGEFEQGT	815
Dd	758	lytstpeegsgwtmnsqalelqgsialylpkpaelfigyfpflfkfgavysrqnkfexsja	817
Qy	816	EAREGSSRLVNLALPtGIRFDKSDODATYNTLTGYVDLVRSNDDCTTTRtRISDSW	875
Dd	818	eaatfdgdvlncslpygltirekiseekmfeslslyldvyrknprrstslmvsasw	877
Qy	876	KTFGTNLARQALVLRAGNHFCFNSNEFAFSOFSTELRGSSRNtVNDGAYOYF	928
Dd	878	tslcnlnarqflasagshltlshvelsgaayelrgshahlynvdcgllysf	930
RESULT	10		
ID	W88424	standard: Protein; 930 AA.	
XX	W88424;		
XX	26-APR-1999	(first entry)	
DE	Chlamydia pneumoniae surface exposed protein Omp11.		
KW	Omp11; outer membrane protein 11; surface exposed protein;		
XX	antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.		
OS	Chlamydia pneumoniae.		
PN	W09858953-A2.		
PD	30-DEC-1998		

Query Match	Best Local Similarity	40.4%	Score 1927	DB 20	Length 930
Matches 414	Conservative 165	Mismatches 326	Indels 48	Gaps 17	
1 MKSPKFEFSEFAIFPSM-IAT--ETVDSASFPGNKNGNEFSRESDACTYLE 56					
1 mkpiphkllissltvlpilisiatygdadalsptcisfdgagstlptystdangtnyvl 60					
57 KGNVLTLENIPGTGTAITKSCFNNTKGLDFTTGNGNSLFPOTVADCTVGAAYNSVVDKS 116					
61 sgnyyl-ndagtgaltgcctfctctgdlitfkgysfsvntvdagsnagaa-stdacka 118					
117 TTFIEFSSLSFIASGSSSTTTCKGAVSOSCTGSLSTIKNVSLIFSKNFST--NGCATTA 173					
119 lltfifsnlsliaaegltvaasqsktlis-sagailtdhgtllifsgvnsneannngallt 177					
174 KTLSTGTMTASLFEENTSKKGAKIQTSDALITINGNGEVSFSDNTSSDSGALFTEAS 233					
178 klslsigtssltisfsnaekrigylissaaslsigsgqlifmnhketggalgafcas 237					
234 VTISSNAKYSFIDNKVYTGASSSTTDDMGGALCAVKTSTDTKVTITLGNOMLFSNNSTT 293					
238 ssaitnssliffignatdaag-----kysalcycktecpcltligskslitfaenssvt 291					
294 AGCAIYVKKLEIASGLTLEFSRNSVNGGTAPKGAIAIEDSGELSLASDGDIVFLGNTV 353					
292 yggaicahgldisaagpfltsmrcnctaagckggaiaiaadsgslsangsdlitlgnlt 351					
354 TSTT-PGTRRSIDIGTSAKMTALRSAAGRAIYFDPV--TTGSSVTVDVLYKNETPA 409					
352 tsstpsltnrliygssaklnlhaaagsgslfcdplasnlttgs---dvltnngps 407					

Db 755 ytgapkgesswyndgcgalelalsiphtalsheglfhayfifkxweasyihbdsfkerntt 814
 QY 817 -AREFGSSRLVNLALPIGIRFDKESDCODATYNTLGYTVDLVNSNPDCITTLRISGDSW 875
 Db 815 lvsifdsdglivnvspigltferfsernerasyeatviyvadvyrknpdcttalllnntsw 874
 QY 876 KTFGTNLARQALVLRAGNHFCFNSFEAFSQPSFELRGSNNYVNDGAKQF 928
 Db 875 ktgtnlstrqgigraglifaifsplevtnlsmeltrgsrsynadlgykltgf 927

RESULT 12
 Y35060
 ID Y35060 standard; Protein; 949 AA.

AC Y35060;
 DT 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae cellular envelope protein.

XX Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

PN W09927105-A2.

PD 03-JUN-1999.

PE 20-NOV-1998: 98WO-IB01890.

XX 04-NOV-1998: 98US-0107078.

PR 21-NOV-1997: 97FR-0014673.

XX (GENSET) GENSET.

PI Griffiths R;

DR WPI: 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX Page 947-949; Disclosure: 1912pp; English.

CC Y34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX Sequence 949 AA;

SO Query Match 38.9%; Score 1856; DB 20; Length 949;

Best Local Similarity 42.7%; Pred. No. 7.1e-115;

Matches 405; Conservative 171; Mismatches 331; Indels 42; Gaps 20;

QY 1 MKSSPKFVSTFAIFP--LSMATEVVLDSASFDGN-KKGNFVSRESQDCACTTLFK 57

Db 22 mktslpwlvsavlafschlqslaneellspddsfngnidsgftfpkts---atyslt 77

QY 58 GNVLTLENIPGTGTAITKSCFNNTKGDLTFTNGNSILFQTVADAGTVGAAYNSVVDKST 117

Db 78 gdvffye-pgkgtplscscfktcdnltflgngnslftgfidagcthnagaa--sttanhl 135

QY 118 TFIQSSLSLTASPGSSITTGKCAVSCSTGSLSTKAVSLFESKNSTNDGCAITAKTLS 177
 Db 136 ltfsgfslstidspsltvtctggtls--sagvnlentlrklvngnfstaagalkgasfl 194
 QY 178 LQGTSMALSENNSSKKGAIOTSDALITTGNOGEVFSFDNMSDPSGAIITFEASVTIS 237
 Db 195 ltysgaallsmssckkgaiatlagariantqxrflsnlstaesgaiddegslis 254
 QY 238 NNAKVSFIDNRKVTGASSTTGDMSSGGAICAYKTSTDRTKVLITGNOMLFSSNNTTPAGA 297
 Db 255 nntkflyf-----egnaektt-----ggaicntkasgspellismnktlffsnvaetsga 305

QY 298 IYAKKLELASGGLTFERNVNGTAKGGAIAEDSGELSLASDSGLYFLGNTVYST- 356

Db 306 lhakklaissggtfeelfrmvssat-pkgsaidsaagselisaetgnlftvrnltltg 364

QY 357 -TPQNRSSIDLTGSAKMTALRSAGRAIFYPIPTGSSSTVTVDVYKVEPTADSALOY 415

Db 365 stdtkrtnainisngkfteliraaknhtlffypdlt--segtsdvtkingagaalpy 422

QY 416 TGNITFGEKLETEAADSNNLTSKLLQVYTLGGTSLKHGVTLOATQOADSRLKM 475

Db 423 qgtllfsgeltadelkvadnlksstfqpvsisgklllkqyvtlestsfsgaagslgm 482

QY 476 DVGTTLE-PADTSTINVLVINISSIDGAKKAKIETKATSKNLKSGITLDPRTGFEN 534

Db 483 dsqctlstagsititnlnvdsiglkpvsitakyaasnkvlsyglnlldlegniyes 542

QY 535 HSLRNPOSYDILELKASGTVTS---TAVPDPIMEGEFHYGQGTWGPVWGTGASTT- 589

Db 543 hmfshdqlfelklktvdadvdtnvdisslipyaeodpenseyfgqgwn-vnwttdeatnc 601

QY 590 --ATFNWTKTGYPNPERISLVPNSLIMNAFIDISLHYMETRANBSLOGDRAFWCAGLS 647

Db 602 kealatwktgtyvpsperkcalvcnltwgyftldrsigqlveigatcmehkqgfwvssmt 661

QY 648 NFEHKDSTKTRGRFRLHLSGCVYVIGNLTCTSDKILSAFQOLFGRDRDYVANKOGTVYG 707

Db 662 nfhktgdenrkqftrhsggyvlgssahpdkdftcfafchlfarddciaimnartlyg 721

QY 708 GTLYVOHNET----YISL-PCKLRPCSLSYVPEIPLVFSGLNSTYTHNDKTKYTTY 761

Db 722 gtlffkshltqpnlylrigrakfseesalekfprelpladvgsfshsdhnmethylsl 781

QY 762 PTVKSGWGNSPALFEGGRAPICL-DESALEBOYMPMKLQFYAHQEGKEGTEAREF 820

Db 782 peesegswneclsgglgldlplfvlsnphlftkftipmkvemyvsqnsfessdgrgtf 841

QY 821 GSSRLVNLALPIGIRFDKESDCODA--TYNLTGYTVDLVNSNPDCITTLRISGDSWKTFG 879

Db 842 sigrlnlslpvgakf-vgqdlgdslydylsgfsvdvynnpqstactllwmspdswklyg 900

QY 880 TNLARQALVLRAGNHFCFNSFEAFSQPSFELRGSNNYVNDGAKQF 928

Db 901 gnlstrqgallrgsnmyynsnclfghyamelrgssrnyvndvgtklrf 949

RESULT 13

W88417 ID W88417 standard; Protein; 928 AA.

AC W88417;

DT 26-APR-1999 (first entry)

XX Chlamydia pneumoniae surface exposed protein Omp4.

DE Omp4; outer membrane protein 4; surface exposed protein; antigen;

KW infection; diagnosis; vaccine; atherosclerosis; asthma.

XX Chlamydia pneumoniae.

PN W0958953-A2.

xx 30-DEC-1998.
 PD
 xx
 xx 19-JUN-1998; 98WO-DK00266.
 PF
 xx
 xx 23-JUN-1997; 97DK-0000744.
 PR
 xx
 xx
 PA (BIRK/) BIRKELUND S.
 FA (CHR1/) CHRISTIANSEN G.
 xx
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygdin P;
 xx
 xx WPI: 1999-105610/09.
 DR N-PSDB; X06816.
 DR
 xx
 xx
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 xx
 xx
 xx Claim 7; Page 40-42; 115pp: English.

CC This polypeptide comprises the novel 98.9 kDa surface exposed
CC protein Omp4 of the human respiratory pathogen *Chlamydia*
CC pneumoniae. Its amino acid sequence was deduced from DNA (see
CC X06816) isolated from a *C. pneumoniae* expression library. The
CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC (see W88417-78), and nucleic acid sequences encoding them (see
CC X06816-27). A new species specific test is claimed that is used
CC to identify mammals (including humans) infected with *Chlamydia*
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of *C. pneumoniae* infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting *in vivo* expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with *C. pneumoniae*.

50 Sequence 928 AA;

Query Match	38.9%	Score 1855;	DB 20;	Length 928;
Best Local Similarity	42.7%;	Pred. No. 8e-115;		
Matches 405; Conservative	171;	Mismatches 331;	Indels 42;	Gaps 20;

[illegible]

Db	344	stdclpkrainlmgsgngkfcteleraaknbnhlffifypdlr--segtssdvlkinnsgalnpy	401
Qy	416	TCNITFTGEKLESEFPADSKULTKLDPVTLSSGTTSLKHGVTLQTAFTQADSLFEM	475
Db	402	qgltlfsgceltladclkvadchlkssftcpvslsgsklllqkyvleatstsgaegslilgm	461
Qy	476	DVGTTLE-PADSTFNNINVINISSIDGAKKAKIEFTKATSKULTLSTGTTLLDPGTFFEN	534
Db	462	dsgcttclstagsltctnnginvdslylqkpvslakasnkvaysgklnliddlegnlYES	521
Qy	535	HSLRNPQSDILELKASCTVTS---TAVNPDPIMGEKPHYGOGTWGPVWGTGASTT-	589
Db	522	hmfsndqjflskltvdsdvgtndvdisallpvpaeedpnssegffggyvn-vnwtldtatnt	580
Qy	590	-ATPNMWTGTYINPPEPISGLVNSLTMNMFIDISSHYMETANEGLOGDRAPFWCGLS	647
Db	581	keatclwtcktfvpspektalvcnclmwyftdlrslsqglveigaqlmekhkgdlwvssmt	640
Qy	648	NFEHAKDSTKTRGRFPHLSGGVYVIGNLTHTCSDKTLASAFCQLGGRDRDYFAVAKNOGTVYG	707
Db	641	nflhktgdenikgfhfhsygyvigsahtpkdpdlftafachlfardckcfiahnnstlyg	700
Qy	708	GLTYQNHET-----YLSL-PCKLRPCSLSVPEIEFVLSGNSLSTHTNDLKTXTTY	761
Db	701	gltlfkfshctlqpnpnylrlgrakfsesalekfprelpdlaldvqvsfshsdnrmethlysl	760
Qy	762	PTVKGSGWGNDSFALFEGGRAPICL-DESALEFEOIMPMLQOFVYAHOGFKEGTEAREF	820
Db	761	pesegswneclaaigldlfpvlsnphplfctklpdmkvemmyvsgnsffessdgrtf	820
Qy	821	GSSRLVNLALPIGIRFDEKSDQDA-TYNTLTGTYVDLVRSNPDCTTTLRISGDSWKTFG	879
Db	821	sigrlrlnlslpvgakf-vqgldigslydydisgffvsdvylnnpqstatclwmspdswkrlg	879
Qy	880	TNLRLQALVTLRAGNHPCFNSFEAFESQSFELRGSSSNVAVDLCARYQF	928
Db	880	gslrtqallrgsnnyvynsucelqfghyamelrlgssrtnyvdvqkrlrf	928
RESULT 14			
Y69368	ID	Y69368 standard; Protein; 945 AA.	
XX	AC	Y69368;	
XX	DT	19-JUN-2000 (first entry)	
XX	DE	Amino acid sequence of the CPN100394 polypeptide.	
XX	KW	CPN100394; Chlamydia infection; immune response; vaccine.	
XX	OS	Chlamydia pneumoniae.	
XX	FT	Key	
XX	FT	Peptide	
XX	FT	Location/Qualifiers	
XX	FT	1..43	
XX	FT	/note="signal peptide"	
XX	PN	WO200011183-A2.	
XX	PD	02-MAR-2000.	
XX	PE	18-AUG-1999; 99MO-IB01449.	
XX	PR	20-AUG-1998; 98US-0097187.	
XX	PR	20-AUG-1998; 98US-0097188.	
XX	PR	20-AUG-1998; 98US-0097189.	
XX	PR	20-AUG-1998; 98US-0097190.	
XX	PR	20-AUG-1998; 98US-0097195.	
XX	PR	20-AUG-1998; 98US-0097196.	
XX	PR	20-AUG-1998; 98US-0097197.	
XX	PR	27-AUG-1998; 98US-0097191.	
XX	PR	17-AUG-1999; 99US-0376770.	

Fri Nov 24 13:49:56 2000

us-09-428-122-2.rag

Page 15

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 08:19:06 ; Search time 35.79 Seconds

(Without alignments)
434.601 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774

Sequence: 1 MKSSPRKVFSTFAIFPLSM.....FELRGSSRNMYNDLGAQYF 928

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCRTUS_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.5	5.7	1338	2	US-08-728-470-9
2	271.5	5.7	1599	2	US-08-617-697-9
3	267	5.6	1529	2	US-08-728-470-10
4	266	5.6	1600	2	US-08-617-697-10
5	244.5	5.1	1026	2	US-08-614-377A-7
6	240.5	5.0	1536	1	US-08-038-682-2
7	240.5	5.0	1536	1	US-08-302-832-2
8	240.5	5.0	1536	2	US-08-302-198-2
9	240.5	5.0	1536	2	US-08-469-880-2
10	240.5	5.0	1536	2	US-08-728-470-2
11	240.5	5.0	1026	1	US-08-617-697-2
12	238.5	4.9	1612	1	US-08-194-290-7
13	234	4.9	1612	1	US-08-169-927-2
14	223.5	4.7	1477	1	US-08-038-682-4
15	223.5	4.7	1477	1	US-08-302-832-4
16	223.5	4.7	1477	2	US-08-302-198-4
17	223.5	4.7	1477	2	US-08-469-880-4
18	223.5	4.7	1477	2	US-08-728-470-4
19	223.5	4.7	1477	2	US-08-617-697-4
20	202.5	4.2	674	1	US-08-317-522A-3
21	202.5	4.2	674	1	US-08-439-818A-3
22	202.5	4.2	674	2	US-08-751-965-3
23	202.5	4.2	674	2	US-08-738-975-3
24	202.5	4.2	674	2	US-08-728-626-3
25	202.5	4.2	674	3	US-08-808-599A-3
26	202	4.2	1912	1	US-08-409-995-4
27	202	4.2	1912	3	US-08-685-467-4
28	201.5	4.2	1160	3	US-08-808-599A-24

29	200.5	4.2	749	1	US-08-317-522A-2	Sequence 2, App1
30	200.5	4.2	749	1	US-08-439-818A-2	Sequence 2, App1
31	200.5	4.2	749	2	US-08-751-965-2	Sequence 2, App1
32	200.5	4.2	749	2	US-08-738-975-2	Sequence 2, App1
33	200.5	4.2	749	3	US-08-808-599A-2	Sequence 2, App1
34	200.5	4.2	749	3	US-08-808-599A-2	Sequence 2, App1
35	181.5	3.8	1248	2	US-08-348-353-17	Sequence 17, App1
36	181.5	3.8	1248	2	US-08-465-965-17	Sequence 17, App1
37	181.5	3.8	1248	3	US-08-465-965-17	Sequence 17, App1
38	169.5	3.6	983	2	US-08-164-292B-26	Sequence 26, App1
39	169.5	3.6	983	3	US-08-845-623-26	Sequence 26, App1
40	169.5	3.6	983	3	US-08-815-927-26	Sequence 26, App1
41	164.5	3.4	1222	2	US-08-682-517-15	Sequence 15, App1
42	164.5	3.4	1252	2	US-08-682-517-9	Sequence 9, App1
43	162	3.4	1394	4	PCT-US95-10661A-2	Sequence 4, App1
44	160.5	3.4	666	2	US-08-737-716-14	Sequence 14, App1
45	159	3.3	907	3	US-08-783-774-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkslesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-9

[illegible]

```

: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1599 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-617-657-9

```

Query Match	Similarity	5.7%	Score 271.5	DB 2:	Length 1599:
Best Local	Similarity	22.5%	Pred. No. 2.1e-12:		
Matches 146:	Conservative	81:	Mismatches 216:	Indels 205:	Gaps 29:
QY	23	TETVLDSSASFQDNKNGNFVRES-----	QEDAGTYY--LFGKGVTLLENI	65	
DB	912	TYTETVAVGASFPDNNGASNTISIRGAKFKFDINTSSLNTITTSDDTYRILIKGNIS--NK		969	
QY	66	PQTGTAITMSC-----FNNTGDLTFEGNGNSLFLQ-TVDDGTVAGAAVNSSVYDK		115	
DB	970	SDDLNIIDKKSDAEIOLGCGNISQCKEGBNLITSSDKYNIITQITIKACVEGGRDSSLAENA		1029	
QY	116	STT-----FLGSSLSIFASPGSSITTKGKAVSCSTGSLTK-----		153	
DB	1030	NLIQIKRELKAGDLNIGSEFNKAETAKKNGSLDITIG---NAGSGVADAKKVTFFDKVDS		1085	
QY	154	-----NWSLFKSKNFSEIDNGCAITAKLLSLGTITMSA--LFSNNTSSKK-----G		196	
DB	1086	KISTDGHNTL--NSEVKTISNGSS--NAGDNSTGLTISAKDVTYVNNVNTSHKITINISAAA		1142	
QY	197	GAIQTSDALITIGNOGEVSEFSDNTSSDSCAAITFEASVYISNNAKVSFLDNKVTGASSST		256	
DB	1143	GNVTTKEGTTINATTSVEVTAQNGIKGN--ITSGNVATVATENLVTTENAVINATSGH		1200	
QY	257	-----TGDMSGCAITAYITSTDTKYTLTGNOMLLESNNT-----STTAGGA		297	
DB	1201	VNISTKTGDIKG--TESTSGVNIITTAGSN--TLKVSNTIGODVITYADAGALTTTAGST		1256	
QY	298	IYVK-----KLELASGGLTFEKSNSVNGGTAPKGAIAI--EDSGEL----		337	
DB	1257	ISATIGNANITTKTGIDINKVESSSSSVTLVAAGCALVAGNISGNVTITADSGKLTSTV		1316	
QY	338	-----SLASDGDV--VFLGNTVSTTGTNRSSLDLTSKAKMALKRAAARAIY		385	
DB	1317	GSTINGTNSVATTSOSGDIEGLTISGNTVNT--ASTGDITIGNSAKVEKKNGA-----		1368	
QY	386	FVPIITGSSITVTVDLKVNETPADSALOYTGNIIFTEGKLESETEAADSKNLTSKLOPV		445	
DB	1369	-----TLAESGKLTITQGGSI		1385	
QY	446	TLSGG--TSLKH-----GVTLQLO-AFTQOADSLEMDVGT-TLEPAD-----T		466	
DB	1386	TSNGCGTTLTAKSSIAGINANAVLNTGLTITGDSKINATSGLITINARDAKLGA		1445	
QY	487	STINNLYINISSIDGAKKAKIETPKATSKNLTLSGTLITDPTGEYEN		534	
DB	1446	ASQDRIYVNTATNAGSGNV---TAKTSSSVNITGDLTINGLINISFN		1490	

RESULT 3
 US-08-728-470-10
 : Sequence 10, Application US/08/728470
 : Patent No. 5928651
 : GENERAL INFORMATION:
 : APPLICANT: Barenkamp, Stephen J
 : TITLE OF INVENTION: High Molecular Weight Surface Proteins
 : TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Shoemaker and Mattare, Ltd.
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 : STREET: Bldg. 1
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: U.S.A.
 : ZIP: 22202-0286
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/728,470
 : FILING DATE:
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/302,832
 : FILING DATE: 16-MAR-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US PCT/US93/02166
 : FILING DATE: 16-MAR-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: GB 9205704.1
 : FILING DATE: 16-MAR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Berkstresser, Jerry W
 : REGISTRATION NUMBER: 22,651
 : REFERENCE/DOCKET NUMBER: 1038-633
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 415-0810
 : TELEFAX: (703) 415-0813
 : INFORMATION FOR SEO ID NO.: 10:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1529 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : US-08-728-470-10

	Query Match	5.6%	Score 267	DB 2:	Length 1529;
	Best Local Similarity	22.6%	Pred. No. 4.5e-12;		
	Matches 146;	Conservative 76;	Mismatches 215;	IndeIs 210;	Gaps 30.
QY	27 LDDSAFPGDKNGKGNFSVES---	OEDAGTYTLFPKGAVYL--ENIEPGTGAIIFKSCFFNNMK	81		
	: : : :	:	:	:	:
Db	846 LNITGTFPNNGTANININOGVKKLOGDINN---	KGLNITTNAAGTQXTIINGINTNEK	901		
	: : : :	:	:	:	:
QY	82 GDLTT-----GN-----GNSLFO-----	IYDACTVAGAAYSSVDKS	116		
	: : : :	:	:	:	:
Db	902 GDLLIKNIKADAEIOIGGISQKEENLNTISSDKVNITNQITFKAGEVGSRSDSSAEMANN		961		
	: : : :	:	:	:	:
QY	117 TT-----FIQSLSLFASPOSSITTKGANVCSTGSLSLK-----		153		
	: : : :	:	:	:	:
Db	962 LTIQTEKLKLAGDLNISGNKAETAKKNSSDLTIG---NASGGNADAKAVTFDYVKDSK		1017		
	: : : :	:	:	:	:
QY	154 -----NVSLFESKNFSTDNGGAIFTAKTLSTLGTMASA-----LPESNTSSRK-----CG		197		
	: : : :	:	:	:	:
Db	1018 ISTDGHNHTL--NSEVKYSNGSS--NAGNDNNSGLTIISADIVYNNNVTSFKRTINISAAAG		1074		
	: : : :	:	:	:	:
QY	198 ALOISDALITFGNOGEVSFSQNTSSDSCGAIFTEASVTIISNAKVSFLDNKVATGASSST-		256		
	: : : :	:	:	:	:

Db	1075	NVTRKEGTTINATGGSVEVTAQNGTILKGN--ITSQVNVVTAATENLVTEMNAVINTSTV	1122
Qy	257	-----TQDMGGALCAIKAKTSTDYKVLTLTGNOMLISNNP-----SITAGCAI	298
Db	1133	NISRTKTDGDKG---IESTSGNNVITASN-FLKVSNIITGQDVTATPAGALTTTAGSTI	1188
Qy	299	YVK-----KELASGGLTFEFSRNVSGCAPRGAAIAI-EDSGEL-----	337
Db	1189	SATTGNAANTTKTGDIENGKEVSSGGVTLVATGATLAVGNISGNTVTTTADSGKLTSTVG	1248
Qy	338	-----SLSDSGDI-VFLGNTVSTTPGTRNSSIDLTGSAKMTALSAAGRAIFY	386
Db	1249	STINGTNSVYTTSSQSGDIEETISGNTVNT---ASIGDILLTIGNSAKVEKKNGAA-----	1299
Qy	387	YDPIITGSSFTTVTVLKVNETPADSALOYTGNIITFGEKLESTEADSKNLTSLKLPVY	446
Db	1300	-----TLAEGSKLTYTGSSIT	1317
Qy	447	LSGG--FLSLKH-----GVTLQIQ-AFTQQADSLEMDVGT-FLERPD-----TS	487
Db	1318	SSNQOTLLTKAKDSSLAGNINAAVNTLNTGTTLTTGDSKINATSGTLTINAKDAKLGGAA	1377
Qy	488	TINLVINISIDGAKKAKIETAKTSKNLTLSGTLTLPDPTGTFEYN	534
Db	1378	SGDITVVNATNAGSGCNV---TAKTSSSVNIITGDLTTLINGLIINISEN	1421

```

1      RESULT      4
2      US-08-617-697-10
3      Sequence 10 Application US/08617697
4      Patent No. 5977336
5      GENERAL INFORMATION:
6      APPLICANT: Batekcamp, Stephen J
7      TITLE OF INVENTION: High Molecular Weight Surface Proteins
8      TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
9      NUMBER OF SEQUENCES: 11
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Shemmaker and Mattare, Ltd.
12     STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
13     STREET: Bldg. 1
14     CITY: Arlington
15     STATE: Virginia
16     COUNTRY: U.S.A.
17     ZIP: 22202-0286
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.30
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/617,697
25     FILING DATE: 01-APR-1996
26     CLASSIFICATION: 424
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER: US 08/302,832
29     FILING DATE: 05-OCT-1994
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER: US PCT/US93/02166
32     FILING DATE: 16-MAR-1993
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Berkstresser, Jerry W
35     REGISTRATION NUMBER: 22,651
36     REFERENCE/DOCKET NUMBER: 1038-557
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: (703) 415-0810
39     TELEFAX: (703) 415-0813
40     INFORMATION FOR SEQ ID NO: 10:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH: 1600 amino acids
43     TYPE: amino acid
44     STRANDEDNESS: single
45     TOPOLOGY: linear
46     US-08-617-697-10

```

Query Match	5.6%;	Score 266;	DB 2;	Length 1600;
Best Local Similarity	22.6%;	Pred. No. 5.7e-12;		
Matches 147; Conservative	75;	Mismatches 215;	Indels 214;	Gaps 29;

QY	31	ASFGNKNKNSVSESOEDGTYTL-FKGVLT-----ENPGTATATKSCF	77
Db	908	STFGESMDNLNITGTTNNGTANINIKGVYKAGDINNKGGALNTTMASGTOKTIIINGNI	967
QY	78	NNFKGDLFT-----GN-----GNSLLQ-----TYDACTVGAANVS	112
Db	968	TNEGDLINIKIKADAEIQGNIISOKEGNLTITSSDKVNITNIIITIKAGVEGGRSDSPA	1027
QY	113	VDRKST-----FIGSSLSFIASPOSSITTGKAVSCSTGSLTK-----	153
Db	1028	ENALLTIQTKEKLAGDLNISGENKAEITAKNSDITIG-----MAGSNAADAKVTPDKV	1083
QY	154	-----NSSLFKNSTFONGAIFATKTLSTLGTMTSA-----LESENTSKK-----	195
Db	1084	KDKISITDGHVLT--NSEVKTSSNGS--NAGDNSTGLIISAKDYVYNNVTSKHTINIS	1140
QY	196	--GGAIGTSDALITTGNOGEVSEFSNDTSSDGAIFTEASVTIISNAKYFIDNKVTGAS	253
Db	1141	AAANVTTKEGTINATNGEVEVTAONGITKGN--ITSONVTATENLVTTENAVINAT	1198
QY	254	SST-----GDMSCGAIKAKYKSTDTKVTILGNOMLLESNNT-----STTA	294
Db	1199	SGIVNISTKTGDIKG--IESTSGNVNITIASGN-TLKVSNITIGODVYVTAADAGALLTTTA	1254
QY	295	GGAIYVK-----FLELASGGLTLFHSNVSNGGAPKGAIAI--EDSGEL--	337
Db	1255	GSTISATGTNANITTKTGDIINGKVESSSGSVTLVAIGALTAIVNGISGNVTITADSGCLT	1314
QY	338	-----SLASDSDI--VELGNVTYSTPTGTRRSIDIGTSAKMTALSAAGR	382
Db	1315	STVGSTINGTNSVYTSOSSDIGETISGNVTNVT--ASTGDITIGNSKAPKANGAA--	1369
QY	383	AIYEPDITTGSSATTVYDLKVNETPADSALOYTGNIIFTEKLESTFEADSKNLNTSKLL	442
Db	1370	-----TLTAEGRKLLTYTG	1383
QY	443	QPVTLSCG--TLSIKH-----GVTLQIQ-AFTQOADSRLMDVGT--TLEPAD--	485
Db	1384	SSITSSNGQTTLLTAKOSSIGAINANVILNTGTGLTTGDSKINATSGTILINAKDAKL	1443
QY	486	--TSTINVLINISSIDAKKAKIEKATSKNLTUSGTTILLDPGTGFYEN	534
Db	1444	DGAASGGRTVYNAATNASGSGN--TAKTSVVNITIGDNTINGLNTIISEN	1491

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/614,377A
;   FILING DATE: 12-MAR-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/194,290
;     FILING DATE: 09-FEB-1994
;     CLASSIFICATION: 435
;     APPLICATION NUMBER: US 07/895,367
;     FILING DATE: 09-JUNE-1992
;     CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;   NAME: Tsao, Y. Rocky
;   REGISTRATION NUMBER: 34053
;   REFERENCE/DOCKET NUMBER: 08106/002001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-542-5070
;     TELEFAX: 617-542-8906
;
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1026 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-614-377A-P

```

Query Match	5.1%	Score 244.5;	DB 2;	Length 1026;
Best Local Similarity	24.2%;	Pred. NO. 1.4e-10;		
Matches 179;	Conservative 86;	Mismatches 257;	Indels 217;	Gaps 33;

QY	338	-----SLASDSDI--VFLEGNVTSTPGNRRSIDGTSAKKTALRSAGR	382
Db	1315	STVGSTINGNSVTTSSGSGIEGTISGNTVNT--ASTGDLTIGNAKAEPAKNGAA--	1369
QY	383	AIIFYDPTTSSSTTVTDVKVNETPADSALOYTGNIIFTEBKUSETEADSKNLTSKLL	442
Db	1370	-----TLTASGKLTQTGT	1383
QY	443	QPVLSGS--TLISLKH-----GYLTQTO--AFTOQASRLEMDVGT--TLERAD---	485
Db	1384	SCITSSNSQOTLLTADSSIAGNINAAVNTLTGTTLTTGDSKINATSGTLTINAKDAKL	1443
QY	486	--TSTINNLVNISSIDGAKRAKIETKAKSKNLTSLTITLDDPTGYEN	534
Db	1444	DGAASGDRVTVNATINASSG--TAKTSSSVNTGDLNLTINGNLTISEN	1491
RESULT	5		
	US-08-614-377A-7		
	Sequence 7, Application US/08614377A		
	Patent No. 5976864		
	GENERAL INFORMATION:		
	APPLICANT: Smt, John		
	APPLICANT: Bingle, Wade H.		
	APPLICANT: No. 5976864ellini, John F.		
	TITLE OF INVENTION: EXPRESSION AND SECRETION OF		
	TITLE OF INVENTION: HETEROLOGOUS		
	NUMBER OF SEQUENCES: 12		
	POLYPEPTIDES FROM CAULOBACTER		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: Fish & Richardson PC		
	STREET: 225 Franklin Street		
	CITY: Boston		
	STATE: Massachusetts		
	COUNTRY: USA		
	ZIP: 02110-2804		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: Floppy disk		
	COMPUTER: IBM PC compatible		
QY	12	TFALFPMIAETVLDDSSAFPCGNKNGNSVRESQDADTTLFLFGNVLLENIP-----G	67
Db	269	TFVAGEVAGAAITLTVGTLGSGAGT-----DVLNWQAAAVTLPLPG--YTIISLETMNT	3222
QY	68	TGTAITKSCFNKTKGDLTFTGNGNSLLEQTVADG-----TVAGAAVNSVYDKSTTFI	120
Db	323	SGAIIITLNTSSGVTG--LTAINTNTSGAAQIVTAGAGONLTATTAQAQANNAVVD-----	375
QY	121	GFSSLTSTASPGSSITTGKRAVSCSTGSLTLKNVLSLFKNSTDNGAITKKTLSLTG	160
Db	376	GRANVTAASGTIVSGITTVAGANSAASGVSV-----SVANSSFTTGTGALAVTG	423
QY	181	ITMGLPSEMTSSKKGAIQTSALITITGQNG--EVSFSDNTSSDGAATF--TEASVTI	226
Db	424	GT--AVTVAOTAGAAVNTLTQADVITGSSSTTAYITVTAATAGAIYAGRVNGAVII	481
QY	237	SNNA-----KVSFTDNKVTGA-----SSSTTGDMSG-----GAICAKYKSTDT	274
Db	482	TDSAAASATTAGKIAATVTLGSPGATIDSSALTLTVNLSTGTSLGIGRALTR-----	534
QY	275	KVTLTGOMLFSNNSTTGAGAIYVKKLELASGGLTFSRNSVNGCTAPKG--GAIAIED	333
Db	535	--TPTANTTLTNVNGLTGT--GAITDSEAAADG---FTTINACSTASSTIASVLAAD	586
QY	334	SGELISADSGDIY-----FLGNVYTS-----PPGTRRSSIDIGTSA	371
Db	587	ATTINTISGDARVITTSHTAALTGLITVINSVGATLGAELATGLVFTGGAGROSLILGATT	646
QY	372	KMTALRSAGRAIFYDPIITGSSSTV-----TDVLKN-----	405
Db	647	K--AIYVAG-----DDIVTVSSALLGAGSVNGGDGTDVLVAANNSSFSFADPAFGG	698
QY	406	ET--PADSALOYTGNIIT--FTGEKLTSEEA-----	432
Db	699	ETLRVAGAAAGSHNANGFTALOLGATAGATPTTNVAVNVGLTVLAAPGTTVTLANAT	758
QY	433	--DSKNLTGSKLQPVYLSGTLSTL-----KHGVILQTOA-----FT	466
Db	759	GTSIDVFLT--LSSSALAAGVIALGVETVNIATDTNTTAHVDFLTLOATSAKSIIVT	816

QY 467 QOADSRLMDVGTLEPADSTI-----NNLVINISSIDAKAKIETKATSK 514
DB 817 GNGGLNLTNGTNAVISEFSDASAVGTAPAYTPVSAANTVEVYIRGAGADSLTGSATA 876
QY 515 NLTL-----SGTITLLDPGTGYENSLRNPOSYDLLELKASGVSTAVTDPDPIGGEFHY 571
DB 877 NDIILGAGADTLVYTGCT-----DTFTGGIGADIFDINALGTSTAFVTTIDAAGDKLDL 932
QY 572 GYGCTWGPVWGT-GASTT 589
DB 933 VGISTNGAIDAGAFGAAYT 951

RESULT 6
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 5.0%; Score 240.5; DB 1; Length 1536;
Best Local Similarity 22.6%; Pred. No. 5.3e-10;
Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

QY 20 MATEVLDSSASFDCNKNNGNFVSRESQEDAGTTLFKGVLTLENIPGTATATKSCFN 79
DB 818 MICKGVVAKKNITFEGBG-NITFEGRKAYTE-----IEGVVITNN--NAVVTILGSDFDN 868
QY 80 TKCDLTF-----FTGNGNSLLEPQYVDACTVAG-AAVNSSVVDKSTFTIGSSLSF 127
DB 869 HOKPLTIKRDVLIINSGLNLAGNIV-----NIAGNLVESNANFKALTFTFNVGGL 920
QY 128 IASPG-SSITTGKAVASC-----STGSLSLTKNS-----LLFSKFNSTDNGCATATKLSL 178
DB 921 FDKKNSNISIAAGARFPDIDNSKNLSTTSSSYRTIISGNTKNG-----DLNT 974

QY 179 T--GTTMSALFSENTSSKKGAIQTSDALTT-----GNOGEVFSFSDNTSSGCAIIF 229
DB 975 TNEGSDTEMOIGGDVSOKEKDELTISSDKINITKQITIKAGVDENDSDPATNNANLTIKT 1034
QY 230 TEASVT-----IS--NNAKVSFDNKKVTGASSSTTGMGSAICAYTSDTKYTLGNOM 283
DB 1035 KELKLTODLNTISGFNKAETIYAKDSDTLTGNINMSADGTAKVYFNQVYDKSLISADGHV 1094
QY 284 LIFS-----NNTSTTGAIAIYVKKLEIASGGLTFSRN-SVNGCTAPKGAIAIEDS 334
DB 1095 TLHSKVETSSGNNNTEDSSDN-----NGLIIDIKNYVNNN-----ITSH 1135
QY 335 GELSLSADSGDIYF-IGNTVYTTTPG---TNKSIDLG-----TSAKMTLRSAGRAIYF 386
DB 1136 KAVISATSGEITTKGTITINATNGVEITTAQTSIIGGIESSGSVTLTATGALAV-- 1193
QY 387 YDPIITGSSSTVTVDLVKVNTPADSALOYTGNITFTGEMKSEFADSKNLTSLQPYT 446
DB 1194 --SNISGNVYTVANGALITTLAIGSTIKGTESV-----TSSQSDIGG-----T 1236
QY 447 LSGGTLSLRHGVTLQTAFTQADSRLMDVGTLEPADSTIN-----NLVIN--- 495
DB 1237 ISGCTVEVK-----ATESLITQNSKRIKATGTGEANTSAFTIGTISGNTVWVANAGD 1291
QY 496 -----ISSIDGA-----KRAKIETKATSKNLTLSGTTITLLDPGTGYENSLRNPOS 542
DB 1292 LTVNGAEINATEGAATLTTSKGLTTEASSHTTSKAGVNLSDQSVAGSINAN-- 1348
QY 543 YDILELKASGVTS 556
DB 1349 ---VTLNTGTGTLT 1359

RESULT 7
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US Pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813

QY 543 YDLELKASGTVS 556
Db 1349 ---VTLNLTGTLTT 1359

RESULT 9

US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 5.08; Score 240.5; DB 2; Length 1536;

Best Local Similarity 22.68; Pred. No. 5.3e-10; Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

QY 20 MATEYVLDSSAFDGKNGNFSVRESOEDAGTYLFGKNTLEINIGCTGAIRKSENN 79
Db 818 MIGGIYAKKNITPEEG-NITFGSRKAYTE-----IEGNVTINN--NANTLTGSPFDN 868
QY 80 TKGDLT-----FTGNGSLLEOTVDAGTVAG-AAVNSSVVDKSTTFIGSSLSF 127
Db 869 HOKPLTIKKDVIIINSGNLTAGNIV-----NIAGNLTVESNANPKATINFTFNVGGL 920
QY 128 IASPG-SSTITGKAVSC---STGSLSLTRKNS---LFSKRFSDNGAIIAKTLSL 178
Db 921 FDNKGNISIIAKGAFKPIDNKSNIITTNSSSTYRTIISGNITUNKG-----DLNI 974

QY 179 T--GTTMSALFSENTSKKGAIGTSDALIT-----GNOGEVSFSDNTSSDGAIF 229
Db 975 TNEGSDTEMQIGDVSQKEGNLTISSDKINITKQITIKAGVDENSDDATNNANLTIKT 1034
QY 230 TEASVT-----IS--NNAKVFIDNKVTGASSSTTGDMSSGAICAKYSTPTKYLGNOM 283
Db 1035 KELKLTODLNISGFKNKKEITAKGSDLTIGNTNSADGTNAKKVTFNOVKRSKISADGHV 1094
QY 284 LFEF-----NNTSTTAGCAIYVKKLELASGGLTFESRN-SVNGCTAPKGAIAEDS 334
Db 1095 TLHSKVTSGSNNTTEDSSDN-----NAGLTIDANVYNNN-----ITSH 1135
QY 335 GEJLSLADSGDIYF-IGNVTSTTPG---TNRSSIDG---TSAKYTLRSAGRAIYF 386
Db 1136 KAVISATSGEITTKGTCTTINATGVNEITAGTISIGIESSSGSVTLTATGALAV-- 1193
QY 387 YDPITGSSSTTVDVKVNTPADSALOYGNIIIFGEKISETEADSKLTLQDPVT 446
Db 1194 --SNISGNTVTVANSALITTLGSLTIKTESV-----TTSQSGDIGC-----T 1236
QY 447 LSGGTLKHHGVTLOAFTQADSRLMDVGTLEPADSTIN-----NLVIN--- 495
Db 1237 ISGTVFVK-----ATPSLTQSNKSIKATTEGANVTSATGTIGTISGNTVWANAAGD 1291
QY 496 -----ISSIDGA-----KKAIEFKATSKNULTSGTITLDPCTGFENHSLRNPOS 542
Db 1292 LTWNGAEINATGCAATLTTSKGLTTEASSHTSAGQVYLSAODSGVAGSIANAAN-- 1348
QY 543 YDLELKASGTVS 556
Db 1349 ---VTLNLTGTLTT 1359

RESULT 10

US-08-728-470-2
Sequence 2, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1536 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 US-08-728-470-2

Query Match 5.0%; Score 240.5; DB 2; Length 1536;
 Best Local Similarity 22.6%; Pred. No. 5.3e-10;
 Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

QY 20 MATEVLDSSAFSGDNKNGNSVRESQEDAGTYLFKGNVLTENIPGTGATKSCFNN 79
 DB 818 MGKGIVAKKNITFEFG-NITFGSKRAVE-----IEGNVTINN--NANVTLLIGSDFDN 868
 QY 80 TKGDLT-----FTGNGNSLFFQTVDAQTAVG-AAVNSVVDKSTTFIFGSSLSF 127
 DB 869 HOKPLTIKKDVIINSNLAGNIV-----NINGNLTVESNANFKAITNFTFNVGL 920
 QY 128 IASPG-SSITTGKAVSC---STGSLSLTKNVS---LFLSKNEFTDNGAIIATKLSL 178
 DB 921 FPNKGNNSIIAKGAFKRDIDSKNLSITNNSSTYRTIISGNITNKG-----DLNI 974
 QY 179 T--GTTMSALFSENTSSKKGAIQTSDALIT-----GNQGEVSEFSDNTSSDSGAIF 229
 DB 975 TNEGSDTEMQIGDVQKQEGNLTISDKINITKQITIKAGVDGEENSDATNNAULTIKT 1034
 QY 230 TEASVT---IS--NNAKVSFIDNKVYTGASSSTTGMGGAICAYKSTDIKVTLTGNO 283
 DB 1035 KELKLTQDLNITSGFNKAELTADGSDLTIGNNNSADGTNAKVTFNQVQKSKISADGHV 1094
 QY 284 LIFS-----NNTSTTAGAIYVKKLELASGGLTFESRN-SVNGGTAPKGAIAIEDS 334
 DB 1095 TLHSKVTSGSNNTTEDSDN-----NAGLTIDAKNVTVNN-----ITSH 1135
 QY 335 GELISADSGDIVF-LGNVTVSTPG---TNRSSIDL---TSKMTALRSAGRAIYF 386
 DB 1136 KAVSISATSGELTTKGTITNATGVEITAOGSLGIESSSGSVLTATGALAV-- 1193
 QY 387 YDPIITGSSITVTVLKVNETPADSALOYTGNIIFTGKLISETPEADSKNLSKLQPYT 446
 DB 1194 --SNISGNTVIYVANSAGALTLAGSTIKGTESV-----TTSQSGDIG-----T 1236
 QY 447 LSGGTLISLKHGVTLOQTQAQADSKLEMDVGTLEPADTSTIN-----NLVIN-- 495
 DB 1237 ISGGTVEYK---ATESLITQSNKIKATTEANVTSAITGIGTISGNTVNVANAGD 1291
 QY 496 -----ISSIDG---KRAKIEFKATSKNLTLSGITTLLDPTGTEYEHNSLNP 542
 DB 1292 LTVNGCAETNAEGATLTITSSGKLTTEASSHTTSKAGOVNLSADGSAVAGSINMAN--- 1348
 QY 543 YDILEIKASGIVTS 556
 DB 1349 ---VTLNTGTLLT 1359

RESULT 11
 US-08-617-697-2
 Sequence 2, Application US/08617697
 Patent No. 597336
 GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 597336-Typeable Haemophilus
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 STREET: Bldg. 1
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/617,697
 FILING DATE: 01-APR-1996
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 05-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Berstreser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-557
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1536 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-617-697-2

Query Match 5.0%; Score 240.5; DB 2; Length 1536;
 Best Local Similarity 22.6%; Pred. No. 5.3e-10;
 Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

QY 20 MATEVLDSSAFSGDNKNGNSVRESQEDAGTYLFKGNVLTENIPGTGATKSCFNN 79
 DB 818 MGKGIVAKKNITFEFG-NITFGSKRAVE-----IEGNVTINN--NANVTLLIGSDFDN 868
 QY 80 TKGDLT-----FTGNGNSLFFQTVDAQTAVG-AAVNSVVDKSTTFIFGSSLSF 127
 DB 869 HOKPLTIKKDVIINSNLAGNIV-----NINGNLTVESNANFKAITNFTFNVGL 920
 QY 128 IASPG-SSITTGKAVSC---STGSLSLTKNVS---LFLSKNEFTDNGAIIATKLSL 178
 DB 921 FPNKGNNSIIAKGAFKRDIDSKNLSITNNSSTYRTIISGNITNKG-----DLNI 974
 QY 179 T--GTTMSALFSENTSSKKGAIQTSDALIT-----GNQGEVSEFSDNTSSDSGAIF 229
 DB 975 TNEGSDTEMQIGDVQKQEGNLTISDKINITKQITIKAGVDGEENSDATNNAULTIKT 1034
 QY 230 TEASVT---IS--NNAKVSFIDNKVYTGASSSTTGMGGAICAYKSTDIKVTLTGNO 283
 DB 1035 KELKLTQDLNITSGFNKAELTADGSDLTIGNNNSADGTNAKVTFNQVQKSKISADGHV 1094
 QY 284 LIFS-----NNTSTTAGAIYVKKLELASGGLTFESRN-SVNGGTAPKGAIAIEDS 334
 DB 1095 TLHSKVTSGSNNTTEDSDN-----NAGLTIDAKNVTVNN-----ITSH 1135
 QY 335 GELISADSGDIVF-LGNVTVSTPG---TNRSSIDL---TSKMTALRSAGRAIYF 386
 DB 1136 KAVSISATSGELTTKGTITNATGVEITAOGSLGIESSSGSVLTATGALAV-- 1193
 QY 387 YDPIITGSSITVTVLKVNETPADSALOYTGNIIFTGKLISETPEADSKNLSKLQPYT 446
 DB 1194 --SNISGNTVIYVANSAGALTLAGSTIKGTESV-----TTSQSGDIG-----T 1236

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 09:30:31 ; Search time 14.34 Seconds
(without alignments)
4106.770 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774
Sequence: 1 MKSSFPKFEVSTFAIFPLSM.....FELRGSSRNRYNDLGAQYQF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2058	43.1	928	2	G81591	polymorphic membra
2	1982	41.5	928	2	B72077	polymorphic membra
3	1965	41.2	936	2	C72078	polymorphic outer
4	1964	41.1	936	2	B81591	polymorphic membra
5	1940	40.6	930	2	A81591	polymorphic membra
6	1936	40.6	930	2	D72078	polymorphic outer
7	1855	38.9	928	2	D72077	polymorphic outer
8	1855	38.9	949	2	F81591	polymorphic membra
9	1573	32.9	841	2	E72130	polymorphic membra
10	1444	30.2	1276	2	G81591	polymorphic membra
11	1442.5	30.2	922	2	B72131	polymorphic outer
12	1441.5	30.2	922	2	F81539	polymorphic membra
13	1383.5	29.0	1407	2	B72078	polymorphic outer
14	1377.5	28.9	973	2	F72076	polymorphic outer
15	1377.5	28.9	995	2	C81593	polymorphic membra
16	120.5	23.5	1013	2	G71460	probable outer mem
17	1051	22.0	987	2	H81722	polymorphic membra
18	911	19.1	867	2	F81721	polymorphic membra
19	863	18.1	878	2	B71460	probable outer mem
20	685.5	14.4	1509	2	H72013	polymorphic membra
21	672.5	14.1	978	2	G72076	polymorphic outer
22	667.5	14.0	978	2	B81593	polymorphic membra
23	666	14.0	947	2	D72067	polymorphic membra
24	666	14.0	1723	2	E72067	polymorphic membra
25	665	14.0	1732	2	C81601	polymorphic membra
26	665.5	13.9	946	2	D81594	polymorphic membra
27	665.5	13.9	946	2	C72075	polymorphic outer
28	647	13.6	952	2	D81593	polymorphic membra
29	646	13.5	934	2	A72075	polymorphic outer

30	603.5	12.6	938	2	H72074	polymorphic membra
31	603	12.6	975	2	F71518	hypothetical prote
32	597	12.5	1016	2	H71460	probable outer mem
33	594	12.4	983	2	A81723	polymorphic membra
34	587.5	12.3	986	2	B81675	polymorphic membra
35	576.5	12.1	1531	2	H71468	probable outer mem
36	559	11.7	1751	2	G71518	hypothetical prote
37	549	11.5	976	2	F81722	polymorphic membra
38	548.5	11.5	1672	2	C81675	polymorphic membra
39	544	11.4	1520	2	A81731	polymorphic membra
40	541.5	11.3	964	2	E71460	probable outer mem
41	533.5	11.2	1450	2	D81675	polymorphic membra
42	523	11.0	1770	2	A71517	hypothetical prote
43	490	10.3	1034	2	F71460	probable outer mem
44	475.5	10.0	514	2	E72076	polymorphic membra
45	453	9.5	1025	2	G81722	polymorphic membra

ALIGNMENTS

RESULT 1

G81591

Polymorphic membrane protein G family CP0303 (imported) - Chlamydomonas reinhardtii (s

C/Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii

C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C/Accession: G81591

R/Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke

Nucleic Acids Res. 28, 1397-1406, 2000

A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae Ar39

A/Reference number: A81500, MUID:20150255

A/Accession: G81591

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-928 <REA>

A/Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38160.1; PID:g718

A/Experimental source: strain Ar39, HL cells

A/Genetics:

A/Gene: CP0303

Query Match	43.1%	Score 2058	DB 2	Length 928
Best Local Similarity	46.7%	Pred. No. 5e-104		
Matches 444	Conservative 142	Mismatches 320	Indels 44	Gaps 20
QY	1	MKSSFPKFEV-STFAIF-PLSMATEVLDSSASFDGKN-GNFSVRSOEDAGTTL	55	
DB	1	MKSQFSLVLSSTLACTSCSTVEAATNAENIGPSDFSGTGTGTPKNT-TTGIDY	58	
QY	56	FKGNVLTENIPGRTATKSCFNNTKGLDFTGNGSLQYDAGVAGAAANSSVDK	115	
DB	59	LGTGILTQNL-GDSALTKGFSDFTESLFAKGSLSLINKS-SAEAAI-SVTYDK	115	
QY	116	STTFIGSSLSFLASPGSSITT-GKAVSCSTGSLTNVLSKNSSTNGCAITA	173	
DB	116	NLSLTGSSLSFLAASVITTPSGKAVKCG-GDLFDNNGITLFPQDCENGCAIST	174	
QY	174	KTLSTGTTMSALPSENTSS-KKGAISDPAITGNGEVSFSDNTSSDGAIFT	230	
DB	175	KNLSLKSTGSSISFEKNSATGKGAICATGTVDITNTATPLFSNNIAEAGCAINS	234	
QY	231	EASVTISNNAKVSFIDNKVTCGASSSTGDMGGAICAYKSTPTKVTLTNGOMLFSNNT	290	
DB	235	TGACTTIGNTSLVFSNSVT-ATAG-NGAL-SCDADVTISGQVTSFGNQ	293	
QY	291	STTAGAIYVKLELAS-GGLTLFSRNSVNGCTAPKGAIAIEDSGELSLSDSGDIV	348	
DB	284	AVANGAIAVAKLTLASGGCGGGSFNSNNIVQGTAGNGAISILAEAGCISAEAGITF	343	
QY	349	LGNTVITTP-GTNRSSIDLTGSAKMTALASAGRAIYFDPTTGSSITVTDLKAVNET	407	
DB	344	NGNAIVATTPOTTKRNSIDIGSTAKITLRAISGHSIFDYDPTTANTADSTDTLNKKA	407	

QY	408	PADSLQYKGNIIIFGKEKSETEADSKNLTSLQAPVLSGGLSKHGVLTQOAFQ	467
Db	404	DKGNSTDYSGSVTSEKSEDEKAYADNLITSLQAPVLTAGNVLKRGVLTDRKGFT	463
QY	466	QADSRLENDVGTLEPA-DTSTINNVLVINISSIDGAKKAKIETKATSKNLTISGTTILD	526
Db	464	TAGSSVIMAGTGTTLAKSTEVEVLTGLSIPVDSLGECKRGVIAAASAKNVALSGTILLD	523
QY	527	PCTGTVEHNSLNPPQSYDILTELKASGTVSTAVTPDPIINGEFHNGYGTWPIYW----	582
Db	524	NOGNAYEHNDLCKTODFSVQLSAGTATTVPAVPATPTHTGYGTW-MTWVDDT	582
QY	583	-GTGASTATFNMKTGYIPIPERIGSLVPNSLMAAFIDISSLHYMETANEGLGDRAF	641
Db	583	ASTPTKTKATTLAMVTNGYLPNPEROGPLVPNSLMSFSDIQAGVIERSAVLTLSDRGE	642
QY	642	WCAGLSNFFHKDSTTRRGRFRLSGYIVGMLHHCSDKILSAAACQLFGRPDRDYVAKN	701
Db	643	WAGAAVNFJDKCKKGEKRRYRHKSGGYAIGAAGTCESENLISFACQDLEGSKRDELVAKN	702
QY	702	QCTVVGGLTYOHNETYISLPCKLRPCLSLVYP--TEHPVFSNLSTYHTDNDLTKRY	758
Db	703	HHDYAGATYIGH----IECGSGTIGCLLDKLRPGSMKHPRVLBEQOLAVSHSNDLTKRY	758
QY	759	TTYPTVKGSGWDSVALEFGGRAPICLDESALEFQOTMPMKQEVYAAHOEGCKEDGTEAR	818
Db	759	TAYPEVKGSGWNNANMMIGAASHSAPYELHCFDYAPAIKINTLYIIRDOSISEGKTGR	818
QY	819	EEGSSRLVLAAPICIRFEDKESDCODATYNTLGTVDVRSNPPCOTTLRTISGSMKTE	878
Db	819	SPDSDNLFNLSPICIVKFEKFSDCNDFSYDLTSLVYPDLTIRNDPCTALYVSGASWETV	878
QY	879	GTNLARQALVLRAGNHFCNSNFEEAFSQFSFELRGSSRNRYVNDLCAQIOF	928
Db	879	ANNLARQALQVRAGSHYAFSPMEFVILQGFVEFVRSSRYVNDLGGKIOF	928

RESULT 2

B72077

Polymorphic membrane protein G family CP0306 [imported] - Chlamydomophila pneumoniae (strain C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: B72077; B81592

R:Kaitman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-388, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: B72077

A:Molecule type: DNA

A:Residues: 1-928 <ARN>

A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AAD18591.1; PID:g4376737

A:Experimental source: Strain CML029

R:Read, T.D.; Brnham, R.C.; Shen, C.; Gill, S.R.; Helgelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: B81592

A:Molecule type: DNA

A:Residues: 1-928 <REA>

A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AA38163.1; PID:g7189223

A:Experimental source: Strain AR39, HL cells

A:Genetics:

:Gene: pmp_-9; CP0306

Query Match	41.5%	Score 1982;	DB 2;	Length 928;
Best Local Similarity	45.2%;	Pred. No. 6.6e+100;		
Matches 429; Conservative 150;		Mismatches 327;	Indels 44;	Gaps 15;
1 MKSSPPKVFVSTFAIFPLSM-----IATFVLDDSSASFQNGKNGNFSVRSQEDA-GTT 53				
	:	:	:	:
	:	:	:	:

Db	1	MKSJLHMFLEISSLALPLSLNFSAFAVLEINLOPNTSFS - PCYITTPAQTINA	
Qy	54	YLFKCNVTLNENPGTGAITKSCNNRTKGLDTEFGNGSNLFPOTVADGVAAGAVN	
Db	59	YNLTGDAVITN-AGSPALTAASKRETTGNLSFPGHGHOFLLOMDIG-ANCFETN	
Qy	114	DKSTFIFFSSLSFASPGSSITTKGKGVSCSTOSLTKKVLSPKSNSTDNWGATA	1737
Db	116	NKLJSFSEFVLSLIQT--TAATGTGAIK-STACSIOSNSCYFPCQNTSNDNGALOG	1727
Qy	174	KTLSLGTMTSALPESENTSSKKGALQTSOALVTITGNOGEVFSODNTSDSGAIIETAS	2333
Db	173	SSISLS-LNPNLTIFAKNKATQKQALYSTGATITNNTLNSASFSENEAANNQCAIYEAS	2311
Qy	234	VTISNAVSVFDKKVNGCASSSTTGDMSGAICAYKTSPTKV-TLNGOMLLFSNNTST	2322
Db	232	SFISNKAISTINNSVTATSA-----TGAIYSSSAPKPVULLSDNELNIGTAT	2855
Qy	293	TAGAIYVKKLELASGGLTTFRSNVNGGTAPKGAIAIEDSGELSLSDSGDIVEGNT	3522
Db	286	TSGGAIYTDNVLVSOGGTLTKFNNSAIDTAPLIGALAIADSGLSLALGDDITPEGNT	3455
Qy	383	V---TSTPGNBRSDIG-TSAKMTALRSAGAIIFYOPTITGSSYTTVDLKVNETP	4088
Db	346	VKGASSSQTTNRNSINGNTMAIYOLRASQNGTYFYDPTITSTAALSDALNLMGPD	4055
Qy	409	ADSALOYGNIITFGEKLESTEADSKNLKSLLOPVTLSGGTSLKHGVTLOQAFTQ	4688
Db	406	LGNPNVGGTIVFGEKLESEMAEADNLKSTIQPPLTLAGGQSLKSGVTIVAKSPSQS	4655
Qy	469	ADSLRMDVGTLEPADTSTNNLVINISLIDAKKKKIETKATSKULLSGTITLLDPT	5288
Db	466	PGSLTMDAGTTLTLAGITNINLVLWVDSLKEPKKATLKTATQASQVTLVLSGLSDPS	5255
Qy	529	GTFEENHSLRNPQSDYDILKLAS--GTVTSTAATPDAIMEKEFHYGQGTWGPVWGTGA	5866
Db	526	GNVEDVSWNNPQYFSCSLTLLADDPANIHITLDAADLEKNPIHMVQGWMA-LSMEDI	5844
Qy	587	ST---TAFNNTKTCGYLPNPRISGLVPSNLMAAFIDISSLHLMETANEGLOGDRAFC	6433
Db	585	ATKSKAALTLTKGYNPNERRGTIVANLWGSFVVRISIOOLVAAKVRQSOETRIWC	6444
Qy	644	AGLSNFHKDSTKTRRGFRHLSGVTYGVGNLHHCSDKILAAFCOLGPRGDYFVXANOG	7033
Db	645	EGISNFEHKDSTKINKGFRHSAGYVVGATTTLASDULITAAFCOLGPKRDRHFINKRA	7044
Qy	704	TVYGGTLYYQHNETYISLPCKLRPCLSYVP--TELPVLFSGNLSYHTDNDLTKRYTT	7604
Db	705	SAYASLTLQHIALTSS-----PSLRLRYPSESEQPVLEDAOISYIYSKMKNTYYTD	7588
Qy	761	YPTKSGNGNSPFLERGGAP-ICLDLSALPEOYMPEMLQOVVYAOHEFKROGTE-AR	8188
Db	759	APKGESSYINDGCALLEASSIPHTALSHEGLFHAIYFPFLIVEASIIHODFFKRNTTVLR	8188
Qy	819	FFGSSRLVNLALPIGIRFDESDQDATALNLGTYDVLYRSPNDCPTTTLRIISGDSWKT	8788
Db	819	SFDSGDITNVSPIGITFEFRSRRNERASYTEATYIYADVAYRKMPDCTTALLINNTSWKT	8788
Qy	879	GTNLARQALVLRAGNHCFNSNFPAFSQFSEFLRGSSRNINVOLGAKYOF	9288
Db	879	GTNLSRQAGIRAGIFAFASPNLEVYNSLMEIRJGSSRSYNAOLGGFQOF	9288

RESULT 3
C72078
polymorphic outer membrane protein g family - Chlamydothrix pneumoniace (strain CWL02999)
C:Species: Chlamydothrix pneumoniace, Chlamydothrix pneumoniace
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72078
R:Kattman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, N.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydothrix pneumoniace and C. trachomatis.
#:Reference number: A72000; MUID:99206506

A:Accession: C72078
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-936 <ARN>
 A:Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PID:AAID18589.1; PID:g437672
 A:Experimental source: strain CW029
 C:Genetics:
 A:Gene: pmp_7

Query Match 41.2% Score 1965; DB 2; Length 936;
 Best Local Similarity 43.4%; Pred. No. 56-99;
 Matches 411; Conservative 179; Mismatches 327; Indels 30; Gaps 16;

1 MKSFFKVFSTRAIF-PLSMIATEYVLDSS-ASFDGNKNGNSVRSQOE-DAGTYLTK 57
 1 MKSSVSWLFFSSILPFLSSLSIVAAEVTLDSNNNSYDGSNGTTFVFTSTDAAGTYSLL 60
 58 GNVTLNIPGTGTAITKSCFNNTKGDLTFTGNGNSLLFQVTDAGTAAVANSVVDKST 117
 61 SDVSFQNALGALIPLASGCLLEAGDGLTFQGNQHALKFAFMNASSAGTAASTAAADKNL 120
 118 TFIFFSSLSFIASPGSSIT-TGKGAVSCSTGSLTKNVSLLFSKFNSTDNGAITAKTL 176
 121 LFNDFSRLLSIISCPDLLSPPTGQCALK-SVGNLSLTGNSQIIFTQNFSSDNGVINTKNF 179
 177 SLTGTTMSALFSENT--SSKKGAIQTSALITLITGNOGENVSFSDNTSSDGAIFTEASV 234
 180 LLSGTSOFASFSTRNOAFTGKOGGVYATGTTIENSPIGVFSQNLAKGSGALYSTDNC 239
 235 TISNNAVFSDNKKVGTGASSSTTGDMSGATCAVKTSTDTKVTLTGNOMLFNSNTSTA 294
 240 SITDNFQVITDGNAMAWAQAQ-----OGGAICC--TTTDTKVTLTGKKNLSFTNNALTY 292
 295 GGALIVKRLKELASGGLTLEFRSNVNGTAPK--GGAIAIEDSGELSLSDSGDIVFLGNT 352
 293 GGATISGLKVSISAGPILFQGN-ISGSSAGOGGGAININASAGELALSATGDTFNNQ 351
 353 VTSTPTNNSIDLGTSAKMTALRSAGRAIYFDPITGSSSTTVYDVKVNETPADSA 412
 352 VTNGSTST-RNAINIIDTAKVTISRAATGOSIIFYDPTNGTAASDTLNLADANSE 410
 413 LOYGNIIFTGKELSETHEADSKNLTKSLDPTVLSGGLSLKKGVTLOTQAFQOADS 472
 411 IEVGAIVFSEKSLPTEKAIANAVSTIRQPAVLARGDLVRCGVYVTFEDLTQSGSR 470
 473 LENDVGTLEPADTS-TINNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPGTG 531
 471 ILMDGTTLSAKKANLSLNLAVNLSSLDGTNKALTEADKNKISLGTALIDTEGSE 530
 532 YENHSLNPSYDILEK---ASGTVSTAVTPDPIMGKHYGYOGTWGPIYVGTGAST 588
 531 YENHNLKSASTYPLLELTGANGTITLGAALSTLTLOEPETHGYOGNW-QLSMANATSS 589
 589 -TATFNWTKGTGYPNPERIGSLVNSLMAAFIDISLHYMETANEGSLUGDRAFWAGLS 647
 590 KISGINTRGVIYSPERKSNLPLNSLMGNFIDIRSLNOLLETSSSEPERELMISGIA 649
 648 NEFHKDSIKTRRGFRHLISGGVIGCNLHTGSDKTLASAFCOLEGRDPIYVAKNOGTVYG 707
 650 NFEYRDMPTRHGRHSHISGYALGITATTTPAEDDLTFAFCOLFARDNNHITGKHGDTYG 709
 708 GTLYOHNEHYISLPCKL-----RPSLSVPTVEIYVLSGNSLSTHTNDLTKXTTYY 761
 710 ASLTFHTTEGLFDIANFLWGMKATRAPVLSISQIIPLESPAKFSYLTHTDHMTYTTDN 769
 762 PYVSGNDSFALFEGFRAPICIDESALFEQYMPFMKLOFVYAHQGEKFEAREFG 821
 770 SIIGSMRNDAFCADLASLPFVIVSYVLLKVEPEVAKVQVIAHQDQFERNAHEGAFN 829
 822 SSRVLNIALPGLIFDKESDQADATYVNLGTYDVLVRSNDDCTTTIRISGDSKKTGTN 881
 830 KSELINVEIPGVTFERDSSEKGTDLTLMYILDAYRRNKCOTSLIASDANMMAGTIN 889

QY 882 LARQALVLRAQHNFECNSNEFAFSQSFELRGSSRNRYNDVJGAKYOF 928
 Db 890 LARQGSVRAANHFOVNPHEIFQGFAYEVRSRRNYNTNLGSMFCF 926

RESULT 4

polymorphic membrane protein G family CP0308 [imported] - Chlamydia pneumoniae (s
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: B81591
 C:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 R.; Read, T.D.; Brunham, R.C.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salze
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: AB1500; M0ID:20150255

A:Accession: B81591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-936 <REA>
 A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PID:AAF38165.1; PID:g718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0308

Query Match 41.1% Score 1964; DB 2; Length 936;
 Best Local Similarity 43.4%; Pred. No. 6,3e-99;
 Matches 411; Conservative 179; Mismatches 327; Indels 30; Gaps 16;

1 MKSFFKVFSTRAIF-PLSMIATEYVLDSS-ASFDGNKNGNSVRSQOE-DAGTYLTK 57
 1 MKSSVSWLFFSSILPFLSSLSIVAAEVTLDSNNNSYDGSNGTTFVFTSTDAAGTYSLL 60
 58 GNVTLNIPGTGTAITKSCFNNTKGDLTFTGNGNSLLFQVTDAGTAAVANSVVDKST 117
 61 SDVSFQNALGALIPLASGCLLEAGDGLTFQGNQHALKFAFMNASSAGTAASTAAADKNL 120
 118 TFIFFSSLSFIASPGSSIT-TGKGAVSCSTGSLTKNVSLLFSKFNSTDNGAITAKTL 176
 121 LFNDFSRLLSIISCPDLLSPPTGQCALK-SVGNLSLTGNSQIIFTQNFSSDNGVINTKNF 179
 177 SLTGTTMSALFSENT--SSKKGAIQTSALITLITGNOGENVSFSDNTSSDGAIFTEASV 234
 180 LLSGTSOFASFSTRNOAFTGKOGGVYATGTTIENSPIGVFSQNLAKGSGALYSTDNC 239
 235 TISNNAVFSDNKKVGTGASSSTTGDMSGATCAVKTSTDTKVTLTGNOMLFNSNTSTA 294
 240 SITDNFQVITDGNAMAWAQAQ-----OGGAICC--TTTDTKVTLTGKKNLSFTNNALTY 292
 295 GGALIVKRLKELASGGLTLEFRSNVNGTAPK--GGAIAIEDSGELSLSDSGDIVFLGNT 352
 293 GGATISGLKVSISAGPILFQGN-ISGSSAGOGGGAININASAGELALSATGDTFNNQ 351
 353 VTSTPTNNSIDLGTSAKMTALRSAGRAIYFDPITGSSSTTVYDVKVNETPADSA 412
 352 VTNGSTST-RNAINIIDTAKVTISRAATGOSIIFYDPTNGTAASDTLNLADANSE 410
 413 LOYGNIIFTGKELSETHEADSKNLTKSLDPTVLSGGLSLKKGVTLOTQAFQOADS 472
 411 IEVGAIVFSEKSLPTEKAIANAVSTIRQPAVLARGDLVLDGVYVTFEDLTQSGSR 470
 473 LENDVGTLEPADTS-TINNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPGTG 531
 471 ILMDGTTLSAKKANLSLNLAVNLSSLDGTNKALTEADKNKISLGTALIDTEGSE 530
 532 YENHSLNPSYDILEK---ASGTVSTAVTPDPIMGKHYGYOGTWGPIYVGTGAST 588
 531 YENHNLKSASTYPLLELTGANGTITLGAALSTLTLOEPETHGYOGNW-QLSMANATSS 589
 589 -TATFNWTKGTGYPNPERIGSLVNSLMAAFIDISLHYMETANEGSLUGDRAFWAGLS 647

Db	590	KIGSINMTRIGTIPSPERKSNIPLANSLMGNFIDISINOLITKSSGEPFELMLSGIA	649
OY	648	NEFHKDSKTRRGRFHHLSGGYVIGGNLHTCSDKILSAACCOLFGHGRDYPVAKNOGTAYG	707
Db	650	NEFYSDSMTRHGPFHISGSGYALGITATTTPADDOLTFAFCOLFADRDHMITKKNHGTYG	709
OY	708	GLYXROHNETYISLCKL-----RPSLSYVPELTPEVFSGLSYTHDNLTCKKTTY	761
Db	710	ASLYHHHEGLDIDIANFLMGKATRAPWVLSIISOIILPISFDAKSYLHHDNMKTYITDN	769
OY	762	PTVKGSGNDSTALEFFGGRAPICLDESALFEQYMEFMKLQFYAHQEGFEKGTIAREFG	821
Db	770	SLIKSMWRDAPCAIDLGASLPFVIVSYPYLKKEVEFPVKQYIYAAQOEFYERYAGRAFEN	829
OY	822	SRLLVNLALPIREDKESDCODATYNTLIGTYVLVNSNPCCITLIRISGSKMTEFTN	881
Db	830	KSELINVELPIQVTERDSKSEKGYDILTIMYILDAYRNPCCOJSLASDANMMAIGTN	889
OY	882	LARQALVLRAGNHCFNSNEAFSQSFELRGSSFNRYVDLGAKYOF	928
Db	890	LARQSFVRAAHNFQVNPHEMELFGQAFREVSRSSNNYNTNLGSKCF	936

RESULT 5
A81591
polymorphic membrane protein G family CP0307 [imported] - Chlamydophila pneumoniae (strata pneumoniae pneumoniae)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81591
R:Read, T.D.; Brnham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1997-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REA>
A:Cross-references: GB:A8002193; GB:A8002161; NID:g7189234; PIDN:AAF38164.1; PID:g7189234
A:Experimental source: strain AK39, HL cells
C:Genetics:
C:Gene: CP0307

Query Match	40.6%;	Score 1940;	DB 2;	Length 930;
Best Local Similarity	43.7%;	Pred. No. 1.3e-97;		
Matches 416;	Conservative 165;	Mismatches 324;	Indels 48;	Gaps 17;

QY	1	MKSSEPKVEFTFALEFJPSM--IAT--ETVVDSSAFSGNKNNGNSVSEODEGATTLF	56
		: : : : : : : : : : : : : : : : : : : :	
Db	1	MKPIPHKLLISTLVTPILLISLATYGADASLSPDTSFGDAGSGSTTPKSTADANGCTNVL	60
QY	57	KGNVLTENIPGTGTAIITKSCFNNTKGDULTFTNGNSLTFQTVDAGTVAAGVANSVYDKS	116
		: : : : : : : : : : : : : : : : : : : :	
Db	61	SGNVIY--NDACKGTAITGCCCTETGTGDLTFTGKVSFEFNIVDAGSNNGAA--STLAKA	118
QY	117	TTTIGFSSLSFTIAPGSSITTTGKAVSCSTGSLSTKVVLSLFESKNFSTD--NGALITA	173
		: : : : : : : : : : : : : : : : : : : :	
Db	119	LTFGTFSMLSEIAAPGTTVAGSKPTLS--SAGALNLTNDGTLTFEQVANGENANNNGAITA	177
QY	174	KTLSLTGTMTALPSENTSSKKKGCAIQTSDALTTTNGNGEVSFSDNTSSDSGALFTFAS	233
		: : : : : : : : : : : : : : : : : : : :	
Db	178	KTLSTISGNTSSITFTLSSAKRLGCAIYSSAASISGNTGOLVPMNNKNETGGGALGEFAS	237
QY	234	VTISNNAVSEFLDKNVYTGASSSTTGDMSGAICAVKTTDTRKVLTEGNOMLTFENNSTT	293
		: : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	238	SSITTONSSLFESGNTATDAAC-----KGAIIYCEKGTETPTLTISGNKSLTFEENSSVT	291
QY	294	AGCAIYVVKLLASGGLTFESRNSVNGCTAPKAGCAIAIEDSGELSLSDSGDITVFLGTV	353
		: : : : : : : : : : : : : : : : : : : :	
Db	292	QCGAICAHGDLDSANGPTLFESENRCGNTAAGKGAIAIADSGSLSLSANQGDITFLGNTL	351
QY	354	TSTT--PGTNRGSDLGTSAKMTALRSAGAIRAIFYDPI--TTGSGSTVTVDVLVNENETPA	409

[illegible]

RESULT 6
D72078
polymorphic outer membrane protein g family - Chlamydothila pneumoniae (strain CML029
C.Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C.Accession: D72078
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:99206606
A.Accession: D72078
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-930 <ARN>
A.Cross-References: GB:AE001627; GB:AE001363; MID:g4376721; PIDN:AAD18590.1; PID:g437
A.Experimental source: strain CML029
C.Genetics:
C.Gene: pmp_8

[illegible]

Db 119 LTFIFGNSLSTIAAPGATVVASGKSTLS - SAGALNTLTDGCTILFSGNYSNEMANNNGGATTT 177
 Oy 174 KTLSTLTGTMGALSTSENTSSKKGCAIOTSDALITGNOGEVSEFSDNTSSDGAALTFEAS 233
 Db 178 KTLSTISGNTSSITFTSNACKKLGAIVYSSAAASISGNTGOLVFNNKKEGTGGALGEEAS 237
 Oy 234 VTISNNAVSFIDNKVVTGASSSTTGDMSCGICAKVKTSTDRKVTLTQOMLFNSNTEET 293
 Db 238 SSTIONSSLFFBSGNATATDAAG-----KGALICEKTIETPTLITSGKSLTFEANSVY 291
 Oy 294 AGGALIVYKLELASGGLTFESRNSVNGGTAPRGGAIAIEDSGELSDASDGIIVLGNV 353
 Db 292 OGGAICAGLIDLASGAPLTFESNNRCGNTAAGKGAIAIADSGISLTSANOGDITFLGNTL 351
 Oy 354 TSTT-PCGNRSSIDLTGAKMTALRSAGRAIYFDPD---TTGSGTIVTVLYKAVNEPA 409
 Db 352 TSTSPSTIRNAIYIGSSAKITTNRAAGOSITYPEDPIASMTGAS--DVLINODPS 407
 Oy 410 DSALOTYNNIIFTGCELSETEADSKNLTSLKLODPVTLSGGLSLKHGVTLOTQAFITQA 469
 Db 408 NSPLDYSGTIVFSGEKLADDEKADNFTSLIKQPLALASGTLAKGVELDVNGFTOTE 467
 Oy 470 DSRLEMDVGTLEPADISTIN--NLVINISSIDGAKKAKIETKATSKNLTLSGTTITLDP 527
 Db 468 GSTLMOGTKLK-ADTEAISLTKLVDLSALEGKSVSITAGANKKITLTSPLEFDS 526
 Oy 528 TGTFEYENSLRPOSDYLDLELKAS-----GVNSTAVTPDPIMGEKFEHYOGTNG 578
 Db 527 SGNFESHTINOAFTQPLVFTTAATAADIIYIDALITSPVOTPEP-----HYIGQHWE 580
 Oy 579 PIWMTGSAATFENMTKGYIPNBERIGSLVPSNLMAAFIDISLAYMETANELOGD 638
 Db 581 ATMADTSTAKSGTMTWVTGTPNPNERRASVYDPSLMAFMDIRTLQIMTSQANSIYQ 640
 Oy 639 RAFMGAGSNFPHDOSTTRGRFRLSGGYVIGGLHCHSCPKLILSAACOLFGGBDROFV 698
 Db 641 RGLMASGTAENFHKDKSTNOAFRHSYGYLVGGAEDFSENFISVAACOLFGDKDLEI 700
 Oy 699 AKNOGTVYGLTYOHNETIYISLPCKLRPC--SLSYVTEIPVLFSGNLSTYHNDNDLKT 756
 Db 701 VENTSHNYLASLYQHRFAFLGLP---MPSFGSITDMLKDPIILNAQLSVSYTKNDMDT 757
 Oy 757 KYTTPYTKSGMDSFALFEFGRAPICL-DESALEFQIMPFMKLQPYAAHQEGFKDGT 815
 Db 758 RYTSYPEAGSGMWTNNGSLALEGSLALYLPREAPEFQGYFPFLKFOAVYSRQONEKEGA 817
 Oy 816 EAREGSSRLVNLALPIAIFREDKESDCODATYNNLGLTVLVLSNPPCTTLRLSGSW 875
 Db 818 EARAADDODDLVNCISPVIRLEKISEDEKKNFEISLAIIGDYRRKNPISRLSLWYSGASW 877
 Oy 876 KYTFGINARQALVLRAGNHGCFNSNEFAFSQFSEFLRSGSRNRYNDLCAKYOF 928
 Db 878 TSLCKNLARQAFILASAGSHLTLSPHVELSGEAAVELRGSAHYVNDCLRYSEF 930

RESULT 7
 D72077
 polymorphic outer membrane protein g family - Chlamydothia pneumoniae (strain CWL029)
 C:Species: Chlamydothia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: D72077
 R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: D72077
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-928 <ARN>
 A:CROSS-References: GB:AF001628; GB:AF001363; MID:94376730; PIDN:AAD18593.1; PID:9437673
 ;:Experimental source: strain CWL029
 ;:Genetics:

A;Gene: pmp_11

Query Match	38.9%	Score 1855;	DB 2;	Length 928;
Best Local Similarity	42.7%;	Pred. No. 5.1e-93;		
Matches 405;	Conservative 171;	Mismatches 331;	Indels 42;	Gaps 20

QY	1	MKSSPKRVFTSTFAIFP--LSMITEVYLDSSAFDGN-KNGNSVSESDOEDACTYLFK	57
Db	1	MKTSIPWLVSVYAFSCHLOLSLNEELSSDSEFNGNIDSGTTTPKTS-----ATTSLT	56
QY	58	GNVTLENIPTGTATITNSCFNNTKGDLTFGNGNSLLFQTVDACTVAGAAVNSSVDKST	117
Db	57	GDVEFYFE-PGKGTPLDSDSCFQOTDNTDLFLNGHSLTFGFIADAGHAGAAA-STTKANL	114
QY	118	TEIFGSSLSFIASGOSSTTGCKGAVSGSTGSLSTKRVSLLSKNESTDNGATITAKTLS	177
Db	115	TFSGSLSLFSFSDSSTVYTTGGQTLSS-SAGVNLNENIRKLVAAGNEFTSADGALTKGASFL	173
QY	178	LGTGTMALFESENNSSKKGAIQTSDALLTTGNGGEVSEFSDNTSSDGAIIFFEASVYIS	237
Db	174	LGTGSGDALFEFNSSSTKGGAIATTAGARIANNGYRFLSNIASTSGAIIDEGTSILS	233
QY	238	NNAKVSTIDNKVTAASSSTTGDMGGAICAKVSTDTKVTLLTGNOMLFESNNSTTAGCA	297
Db	234	NNKFLYF-----EENAKT-----GGAICNTKAGSCELLISNNKTLFISNAVETSGA	284
QY	298	IYVKLELASGGLFLFSNNSVNGCTARKGAIIDEJGELSLSDSDIYFLGTVST-	356
Db	285	IHAKKLAISSGGLFEFLRNYSAT-FKGAISIDASGELSLSAETINITFVRNTLTTTG	343
QY	357	-TPGTRSSSIDLGTSAKMTALRSAGRAIYFDEITTGSSSTVYTDVKNETPADSALOY	415
Db	344	STDTPKRAAINIGSNGKFTLELRAKNHTIFFEYDPT--SEGTSSDVLIKINGSAGALNPY	401
QY	416	TGNITTFGEKLSFEHDASKTLTKLQPTLSGGLSLKHWGLTQQAFTQADSLEH	475
Db	402	OGTLTFGEITLTADELKADWLKSTFOPVSLSGKLLDKQVLTLESTPSOEKSGSLGM	461
QY	476	DVGTGLE-PADTSTINNLVINISSIDAKRAKIETKASTKNLTSGLTITLLDPTGYEN	534
Db	462	DSGTLSTTAGSITITNLGINVDSLGLKQPVSLAKGASKMVIYSGKLNLDIEGNYES	521
QY	535	HSLRNPOSYDLELKASGYTIS----TAVTPDPLMGKFRHYGVTGMPYWGIGASTT-	589
Db	522	HMFSGDOLFLSLKLTVDADVDTNVDISLPLVPRAEDNSEEGFGQGN-VNMTDTPTNT	580
QY	590	-ATEFNMTKTYINPNERIGSLVNSLMNFAFIDSSHYLMETANEGLQDRAFGCLS	647
Db	561	KEATATATKTFVSPERSKALVNTLMGVPTDRSLQOLVEIGATGMEHKOGFWYSMT	640
QY	648	NFHKDSTKTRGRHLSGGVYVIGNLHTGSDKILLSAFCOLFGRDRDYFAVAKNOGTVYG	707
Db	641	NFLKHTGDENKRKGRHNTSGGVIGSAYHTPEADDLTFAPCHLEFARDKDCFIAHNSTRYG	700
QY	708	GLTVYQNHET-----YISL-PCIKRPGSLSVPRTEIVLSEGNLSYTHNTDNLKTKXTTY	761
Db	701	GLTFEKFHSHTLOPONYTRLGRAKFSESAIEKFEPRLEPLALDVQVSEFSHNSRRETHTSL	760
QY	762	PTVSGMGNDSFALFEFGRAPICL-DESALFEQGMPPMKLOFYVAHDEGKEOGTEAREF	820
Db	761	PESGSMWNEBIAAGIGDLDPVLSNHPJPKFTIIPQMKEMYVVSQNSPFESSDORGF	820
QY	821	GSRLVALALPIGIRFDKESDCODA-TYNLTGLTVLVLSNBPDCITTLAISGDSKATFG	879
Db	821	SIGRLNLSITIVGAKF-VQOIGIGSYTYLDSGEFVSUVYRNNDPSTATLVMSDSKIRG	879
QY	880	TNLRQALVALACGNHCFNFSFEAFSOFSELRSSSNVYVUDLGAKYQF	928
Db	880	GNLSRQAFLLRGSNNYYNNSNCELFHGYAMELDRSSSNVYVUDGTRKRF	928

[illegible]

RESULT 10
C81591
Polymorphic membrane protein G family CP0309 [imported] - Chlamydomophila pneumoniae (strain A639)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81591
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
M.; et al. Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
Nucleic Acids Res. 28: 1937-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: C81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1276 <FEA>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38166.1; PID:g7189222
A:Experimental source: strain AK39, HL cells
C:Genetics:
C:Gene: CP0309

Query Match	30.2%;	Score 1444;	DB 2;	Length 1276;
Best Local Similarity	29.5%;	Pred. No. 1.5e-70;		
Matches 382;	Conservative 157;	Mismatches 370;	Indels 386;	Gaps 26;

Oy	1	MKSSPPKVFSTFAFPL-SMATETVLDDSSAFPGNKGNG--TSVRSDQEDAGTYLFLK	57
		: : : : : :	
		: : : : : :	
Db	1	MKSYLPWLLTSLALFSLHPPLMAANTLDDSSDNYENGSSGSAFAFKTSPASGTYTLT	60
Oy	58	GNVLTEILPGIGTAT---KSCFNNTKGDILFTFGNGSLTFPTVAGIVAGAAVSSVVD	114
Db	61	SDVSTTNN-----SALTPAKDSCTETGALSLFVGADHSLVLOTI-ALTHDGAALNNT--N	113

Oy	115	KSFIKFGSSLSFIASPGSSITIGKAVGSCST-----GISLTKRNVLSLFKRFSPNDNGA	170
Db	114	TALSSGSSSLILDSAPATGSGKGAI-CVYTBEGGTAFETDMAVYLQKNTSEKDGA	172
Oy	171	ITAKTLSTGTMTSALBSEMTSSKKGAIOTSDALITITNOGEVSPSDTSSDSGAIAIF-	229
Db	173	VSAYSIDAKTTTAALLDQNTSTKNGALCSTANTTYVQNSGTVFSSMTATDKGGIYS	232
Oy	230	-----TEASVTISNAKVSFIDNKVTG-ASSSTTGDM	260
Db	233	KEKSDTLDANTGVTFEKSNTAKTGAWSSDDNLALGTNOVFOEKNKTKGSAQAQNNEG	292
Oy	261	SGGAICAV-KTSD------TKVLTGNOMLFSNNTS	291
Db	293	CGGAICCYLATADTKGLAISOMEMSFTSNNTTANGAGAIATKCTLDGNTTLTFDQNTA	352
Oy	292	T-----	292
Db	353	TAGCGGAIYETEDFSLKSGTGTVESTNTAKTGALYKGNSSLTNGNLFFSGNKATG	412
Oy	293	-----	292
Db	413	PSNSSANOEGCGAIIISFLSASVSTKKGMLIEDNENVSLSGNTATVSGAIYATKCALH	472
Oy	293	-----	292
Db	473	GNTTLTFDQNAETAGAIATTEDEFTLIGSTGTVFSTNTAKTAGALHTKNTSFTNKK	532
Oy	293	-----	292
Db	533	ALVESGNSATATATTTDOEGCGAIIICNISSEDIATKSLTLEMESLFFINNTAKRSOG	592
Oy	293	-----TAGGAIVYKLELASGTLFFSNSVNGSTAPKAGAI	329
Db	593	GIYAPKCVISSESIENDGNTAETSGAITSKULSTANGAPVSFTNNS--GG---KGAII	647
Oy	330	AIEDSGELSLADSGDIVFLGNTVSTTPGTNRSIDIGTSAKMTALRSAGRAIYFDP	389
Db	648	YIADSGELSLAIDGDTFGGNATEGTSTPN--SHILGAKAKITKLAAPHTIYFYDP	705
Oy	390	IT-----TGSSITTYND-----VLKVFETPADSALQYTGNIITFGKL	426
Db	706	ITMEPAPASGGITIEVLINPVYKAIIVPPPOKNGPIASVPVPAAPNPMTGTVFSSGCL	765
Oy	427	SETEPADSKNLTSLKLOPVLSGCTLSLKHGYVLQTAFTQOADSRLKEMDVGTTLEPADT	486
Db	766	PSQASIPANTTTILNQKINLWAGNVYKKGANLQYVSTQQPDSTVFEMDACTTLETTT	825
Oy	487	ST-----INLVINISSIDGAKKAKIETKATSKNLTLSGTLTLLDPGTSTFYEHSKLP	540
Db	826	NNTDGSIDLKNLNVNDALDLEGKMRITIAVNVISGELKISGDLKFHNBSGFYDNGGLAN	885
Oy	541	QSYDILELKA-SGTVSTAVTPDPIMGKFFHYGQTCG--PIVMQGSTATFENMKT	597
Db	886	LNLEFLDLSSTSGVNDNDNPIPSMAAPDYGQSSWMLYKVV-GAGKATVILVEMQAL	944
Oy	598	GIYFNPERRIGSLVPSISMAAFIDISSLHYMETANEGLOGDRAFWCAGLSNFFKROSTKT	657
Db	945	GYTFKPELRATILVPSNLSMNVYVNIHSIOEIIATAMADASHPGIIGWIGIGIAFHDOKE	1004
Oy	658	RGRFHLISGGVYIGGNLHTCSDKILSAFOLGGRPDVFAVKNNGTGYGGLYVQNHET	717
Db	1005	NAGFRLLISRGYVIGGSHMTTPOETTFPAVHSQLFGKSKDIDVSDIKSOYIAGSLCWO-SSY	1066
Oy	718	YISLPCKLPRCSLSYV---PREIPVLFGSNTSYTHTDLTKRYTYTPTVYVSGMGNDSF	773
Db	1064	VIPHSSLSRHHVLSKYVPELPGETPLVLYLHGOVSYGGRHNHMTTKLANNQGSMDWSHF	1122
Oy	774	ALFEGGAPICLDESALEFEDYIMPFMKLQYVYAHQBEFKROGTEADAEFGSSRLYNALALIG	833
Db	1124	AVEGGSSLPVDLNNRVL-TSYSPYVKLQVYVNOKEFCQVADPRFPAASHLVANVISIMG	1182

```
OY 834 IRFDKESDCODATYNLTGTYVDLVRSPNDCCTTLRISGDSWKRTGTNLAQOALVLRAGN 893
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1183 LTFKESAKKPEPALLLTGYADAYRDHPHCILSL-TNTQSTMTATINLSRQAFEAASG 1241
OY 894 HFCFNSNEAFSOFSEFLRGSSRNATVDLGAKYQF 928
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1242 HLKLLHGLDCFASGSCELRSSRSRYANAGTRYSF 1276

RESULT 11
B72131
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: B72131
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: B72131
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <RAN>
A:Cross-References: GB:AE001585; GB:AE001363; NID:g4376255; PIDN:AA018463.1; PID:g437626
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_1

Query Match 30.2%; Score 1442.5; DB 2; Length 922;
Best Local Similarity 36.9%; Pred. No. 1.1e-70;
Matches 351; Conservative 156; Mismatches 381; Indels 63; Gaps 26;

OY 5 FP-KVESTFAIFPLSMATEVYLDSSASFDGNKNGNSVRESQEDAGTYLFRGNVYLE 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 FPLVFSFTLLSVFDTSLATITSLTPEDSFHGDSON--AERSYNVQAGDVYSLTGDVSI 65
OY 64 NIPGTGAITKSCFNNTKGLDFTGNGNSLLFQTVDACTVAGAAVNSVVDKSTP-FIGF 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 NV--DNSALNKACFNVTSGVTFAGNNHGLVFNNISGTTKEGAVLCCODPQATARESGF 123
OY 123 SLSLFIASPGSSITTGKAVSC--STGSLSLTKNVSLLFKNSFTDNGAITAKTSLTG 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 STLSFIQSPGRIKEQG-----CLYSKNALMLNNYVRFEDONQSKTKGAGAVITVG 178
OY 181 TTMSALFSEMTSSKKGAIGTSDALITGNGEVSFSDNTSSD--SGAIFTEASVTISNN 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 NYDSVSTFQNAAT-FGGAIHSSGPIQIAVNOAEIRFAONTAKNGSGGALISDGDIDON 237
OY 240 AKVSFIDNKVTGASSSTTGDSGGAICAYKTSTDR---VTLTGNQMLFSSNNTSTTGA 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 AYVLFRENE-----ALTTAIGKGAVCCLPITSSSTPVIPTFSDNKQLVERNHSIMG 292
OY 296 GAIYVKKLELASGGLTIFSRNSVNGGTAPKGAIAIEDSGELISADSGDIYVLGNVTYS 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 GAIYARKLSISSGGPTLFINNISYANSQNLGALAIIDTGEISLSAEKGTITFGQ--RT 350
OY 356 TTTPGNRSIDLGTSAKKTALRSAGRAIYFYPDTTSSSTTVIDVLYNTEPDAISAOY 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 SLPELN--GIHLLONAKLKIQARNGVSIETFDPT--SEADGSIQOLINIDPKKK--EY 404
OY 416 TGNITFTGKLETSEADSKNLTKLQPTVLSGGTSLKHGVTLLQTAFTQOADSRLAM 475
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 TGTILFSEK---SIANDPRDFKSTIPQNVNLSAGYLVIKGAETVSKFQSPESHVL 461
OY 476 DVGTTL-EPADSTINNLVINISSIDGAKKAK-IETKATSKNLLISGTTLLDPTGYE 533
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 DLGKRLIASKEDIAITGAIIDISLSSSTAIVIKANTANKOISVTDIEHISPTGNAE 521
OY 534 NHTSRNPOSYLLELK--ASGTVSTA-----VTPDPIMGKGFHYGYGTGPIYV--CTG 585
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 DLRRNSOTFPLLSLEPGAGSVTVTADPLVPSP-----HYFGQNW-KLAWTGTG 572
```

```
OY 586 ASTATENWTKTGYPINPERIGSLVPSNLSMAFISSLAHYMTANEGLOGDRAPACAG 645
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 -NKGEFFWDKINKPREKEGNLVPLINMGNAVDSLMOVOTTHASSLOTDLGLIDG 631
OY 646 LSNFFHKDSTKTRRGFRHLISGYYIGGLATCSDKILISAFCQLFGDRDVFYAKNOGTV 705
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 IGNEFHVASBDNIIRYRNSGCVLVSNNETTPKHYTSMASFQSFSDKDYAVANNBY 691
OY 706 YCGTLYIOHNET-----YISLPCKLRCSLSYVTELP-VLESGNLSYTHTDLTKY 758
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 692 YLGSILYQYTTSLIGNIFRYASRNPNVNGILSRRELQNPMLIFHFLCAYGATMDKTDY 751
OY 759 TTYPTVKGSGNDSEAFLEFGRAPICDESA-LFEQYMPKLOFVVAHOEFKEQCTEA 817
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 752 ANFMVKNKSMNNOCMAIECGSMPLVFNENRFLQGAIFPKLQLYAYQGFRETTADG 811
OY 818 REFSSRLVNLALPIGIRFDKESDCODATYNLTGTYVDLVRSPNDCCTTLRISGDSWK 877
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 812 RRFNSGSLTISVPLGIRFEKXIALSDVLYDFSFYIPDIFRKPDSCEALVIGSDMLV 871
OY 878 FGTNLARQALVLRAGNHFCFNSNEAFSOFSEFLRGSSRNATVDLGAKYQF 928
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 872 PAHVSRHAFVGSCTGRHENDYTELLCRGSIECRPHARNINCGSKFRP 922

RESULT 12
F81539
polymorphic membrane protein G family CP0770 [imported] - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: F81539
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C: Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, C.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: F81539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <RAN>
A:Cross-References: GB:AE002237; GB:AE002161; NID:g7189664; PIDN:AAF38570.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0770

Query Match 30.2%; Score 1441.5; DB 2; Length 922;
Best Local Similarity 36.9%; Pred. No. 1.3e-70;
Matches 351; Conservative 156; Mismatches 381; Indels 63; Gaps 26;

OY 5 FP-KVESTFAIFPLSMATEVYLDSSASFDGNKNGNSVRESQEDAGTYLFRGNVYLE 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 FPLVFSFTLLSVFDTSLATITSLTPEDSFHGDSON--AERSYNVQAGDVYSLTGDVSI 65
OY 64 NIPGTGAITKSCFNNTKGLDFTGNGNSLLFQTVDACTVAGAAVNSVVDKSTP-FIGF 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 NV--DNSALNKACFNVTSGVTFAGNNHGLVFNNISGTTKEGAVLCCODPQATARESGF 123
OY 123 SLSLFIASPGSSITTGKAVSC--STGSLSLTKNVSLLFKNSFTDNGAITAKTSLTG 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 STLSFIQSPGDIKEQG-----CLYSKNALMLNNYVRFEDONQSKTKGAGISAGAVITVG 178
OY 181 TTMSALFSEMTSSKKGAIGTSDALITGNGEVSFSDNTSSD--SGAIFTEASVTISNN 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 NYDSVSTFQNAAT-FGGAIHSSGPIQIAVNOAEIRFAONTAKNGSGGALISDGDIDON 237
OY 240 AKVSFIDNKVTGASSSTTGDSGGAICAYKTSTDR---VTLTGNQMLFSSNNTSTTGA 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 AYVLFRENE-----ALTTAIGKGAVCCLPITSSSTPVIPTFSDNKQLVERNHSIMG 292
OY 296 GAIYVKKLELASGGLTIFSRNSVNGGTAPKGAIAIEDSGELISADSGDIYVLGNVTYS 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 GAIYARKLSISSGGPTLFINNISYANSQNLGALAIIDTGEISLSAEKGTITFGQ--RT 350
OY 123 SLSLFIASPGSSITTGKAVSC--STGSLSLTKNVSLLFKNSFTDNGAITAKTSLTG 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 STLSFIQSPGDIKEQG-----CLYSKNALMLNNYVRFEDONQSKTKGAGISAGAVITVG 178
OY 181 TTMSALFSEMTSSKKGAIGTSDALITGNGEVSFSDNTSSD--SGAIFTEASVTISNN 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 NYDSVSTFQNAAT-FGGAIHSSGPIQIAVNOAEIRFAONTAKNGSGGALISDGDIDON 237
OY 240 AKVSFIDNKVTGASSSTTGDSGGAICAYKTSTDR---VTLTGNQMLFSSNNTSTTGA 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 AYVLFRENE-----ALTTAIGKGAVCCLPITSSSTPVIPTFSDNKQLVERNHSIMG 292
OY 296 GAIYVKKLELASGGLTIFSRNSVNGGTAPKGAIAIEDSGELISADSGDIYVLGNVTYS 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 GAIYARKLSISSGGPTLFINNISYANSQNLGALAIIDTGEISLSAEKGTITFGQ--RT 350
```


QY 712 YOHNETYISLPCKLRPCSLSYV-----PTEIPVLFSGNLSTYTHDNDLTKTYTPYVKG 767
 Db 1190 AQ-SSVIVPLHSLRKHVLSKVLPELPGETPLVHGQVSYGRHHMTTKLANNTQKSD 1248
 QY 768 WGNDSFALEFGGRAPICIDESALFEQYMPFMKLOFYVAHQEFKEDEGTAREFGSSRLVN 827
 Db 1249 WDSHSAFAVGGSLPVDLNYRYL-TSYSPYKLOQVYVNOQGFQEAADPRIFDASHLVN 1307
 QY 828 LALPTGIRPDKESDCODATYNNLTGTYVDLVRSNPDCTTTLTSGDSKTEGTNLAQAL 887
 Db 1308 VSIPLMGLTFKHSAPSPSALLTLGLADYADYDHPHCLTSL-TNGTSMSTFATNLSROAF 1366
 QY 888 VLRAGNHFCFNSNFEAFSQFELRGSSSRNYVDLGAKYQF 928
 Db 1367 FMAAGSHLKLHGLDCCFAAGSCELSRSSRYANNGTRISF 1407

RESULT 14

F72076
 polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: F72076
 R:Kaitan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: F72076
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-973 <AR>
 A:Cross-references: GB:AE001629; GB:AE001363; NID:g4376734; PIDN:AA18595.1; PID:g437673
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: pmp_13

Query Match 28.9%; Score 1377.5; DB 2; Length 973;
 Best Local Similarity 34.8%; Pred. No. 4.1e-67;
 Matches 350; Conservative 153; Mismatches 391; Indels 113; Gaps 25;

QY 1 MKSSFPKVFST-----FAIFPLSMIAETVLDSSAFDGNKNFVSRESOEDAGTYL 55
 Db 1 MKTSIRKFLISTTLAPCFA-----STAFVEYVIMPESENFDSGSKIFPYTTLSDPGTLICI 56
 QY 56 FKGAVTLENIPGTGTATKSCFNNTKGDLTFTGNGNSLLEQIVDAGTVAGAANSSVYDK 115
 Db 57 FSGDLIYANLDAISRSSSCFSNRAGALQILKGCVSFLNI-RSSADGAATSSVITON 115
 QY 116 ---STTFIGFSSLSF-----IASPGSSITTGKAVSCSTGSLSTKRVSLFSKNFS 164
 Db 116 PELCPLSFGFSQMIFFDNCESLTSOTSASNVIPHASAIYATPML-FTNNDLSILFOYRNS 174
 QY 165 TDNGCAITAKLTLSTGTMSALFSNTSSKKGAIQTSDALITNGOEVSFD----- 218
 Db 175 AGFGAIGKTSITLIENTKSLLEFNGSISNGALTGSAALINLINSAPVIFSTVATGIY 234
 QY 219 -----NTSSDGAIFTEASVTSINNKAVSFIINKVYAGASS- 254
 Db 235 GGAIVLTGSGMISGNLGVLFVNNSSRSGAIIYANGVNTSSNNDLTFQNNATSPONS 294
 QY 235 -----STTGDMG-GGAI-C--AYKTSDDTKRVTLTGNOMLLFSSNNTSTAGAIY 299
 Db 295 PAPTPPPPAVTPPLGLGGAIFCTPPATPPPTGSLTISGENSVTFLENIASSEGGLY 354
 QY 300 VKKLELAGGLTFSRNVNGSTAPKGAIAIEDSGELSLSDSDIYFLGN-YVTSTTP 358
 Db 355 GKKSISIDNKSITFL-----GNTAKGGAIAIPESGELSLANOGDIIFFNNLSITSGTP 409
 QY 359 GTNNSSIDLGTSAKMTALRSAAGRAIYFDPITGG--SSTVTVDLKYVN-ETPADSAIOY 415
 Db 410 --TNRSHFGADAKFAITGATGCTIYTFYDPTLSDLSAASAAATVAVNPKASADGA--Y 465

QY 416 TGNLIFTEGKLTSEPAADSKNLTKLOPVTLLSGTSLSKHGVLTQOAFVQOADSRLM 475
 Db 466 SGTIVFSGEVLTAFAAPPAVATSTLNQKLEEGTTLARAGALNVHNFVODEKSVYIM 525
 QY 476 DVGTLLEPADS-----TINNVLINISIDGAKKAKIEFKATSKNLTLTSGTTLIDPT 528
 Db 526 DAGTLTATNANNNDGAILTKLVINIDSLDGTAKAANVNOSTNGALTTIGTIGLVKNS 585
 QY 529 GTFENHSLRNP--QSDVILEKA-SGTVSTAVTPDPIMEKFEHYGQGTWGPVWGTG 585
 Db 586 QDCCNHMHFMKDDLOQVPILELKAISNVTYTTIDSLGTNGVQSGPYGQGTWEFTIDT- 644
 QY 586 AASTATFMTWGTGYPIPERIGSLVPSNLSMAAFDIDSSLHMLMTANEGLOGDAFACAG 645
 Db 645 -THVTGWKKRTGYLPHPERLAPLIPNSLVANVIDLRVSOASADADEDVPG-KQLSITG 702
 QY 646 LSNEFHKDKTKTRGFRHLSSGYVIGNLHTCSDKILSAFCQLFGDRDVFYAKNOGTV 705
 Db 703 ITNPFHANHTGDARSYRMGGGYLINTYTRITPDALSLGFGQLFTSKOYLVEGHGSHV 762
 QY 706 YGGLTIYOHNETYISLPCKLRPCSLSYPTIPLVFGSNLS--YTHDNDLTKTYTPY 762
 Db 763 YFAVYVSNITKSLFG-----SSRFESGTSRVYTSRSNEKVKTSYTKLP 806
 QY 763 TVKSGWGNDSFALFEFGGRAPICIDESAL-FEQYMPFMKLOFYVAHQEFKEDEGTAREFG 821
 Db 807 KGRCSWMSNMCNLGLEEGMLPTLSSRLNLKQIIPFYKAEVAYVTHGIGDENTPEGHIFG 866
 QY 822 SSRVLNALPIGIRPDKESDCODATYNNLTGTYVDLVRSNPDCTTTLRISGDSKTEFTGN 881
 Db 867 HGHLINAVPAGVYFEGKNSHNRPDFYTLIVAYADYVRHNDPCTTLPINGATWTSIGN 926
 QY 882 LARQALVLRAGNHFCFNSNFEAFSQFELRGSSSRNYVDLGAKYQF 928
 Db 927 LTRSTLLVQASSHRSVNDVLEIFGHCCGDIRKTRISQYTLIDIGSKLP 973

RESULT 15

C81593
 polymorphic membrane protein G family CP0299 [imported] - Chlamydia pneumoniae (s
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81593
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gilly, S.R.; Heidelberg, J.F.; White, O.; Hicke
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150235
 A:Accession: C81593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-995 <REA>
 A:Cross-references: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF8156.1; PID:g718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0299

Query Match 28.9%; Score 1377.5; DB 2; Length 995;
 Best Local Similarity 34.8%; Pred. No. 4.2e-67;
 Matches 350; Conservative 153; Mismatches 391; Indels 113; Gaps 25;

QY 1 MKSSFPKVFST-----FAIFPLSMIAETVLDSSAFDGNKNFVSRESOEDAGTYL 55
 Db 23 MKTSIRKFLISTTLAPCFA-----STAFVEYVIMPESENFDSGSKIFPYTTLSDPGTLICI 78
 QY 56 FKGAVTLENIPGTGTATKSCFNNTKGDLTFTGNGNSLLEQIVDAGTVAGAANSSVYDK 115
 Db 79 FSGDLIYANLDAISRSSSCFSNRAGALQILKGCVSFLNI-RSSADGAATSSVITON 137
 QY 116 ---STTFIGFSSLSF-----IASPGSSITTGKAVSCSTGSLSTKRVSLFSKNFS 164
 Db 138 PELCPLSFGFSQMIFFDNCESLTSOTSASNVIPHASAIYATPML-FTNNDLSILFOYRNS 196

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 11:21:07 ; Search time 42.6 Seconds
(without alignments)
695.955 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774
Sequence: 1 MKSSFPKVFSTFAIFPLSM.....FELGRSSRNRYNDLGAKYOF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	541.5	11.3	964	1	OMPE_CHLTR
2	490	10.3	1034	2	OMPE_CHLTR
3	269	5.6	2249	1	190K_RICRI
4	244.5	5.1	1025	1	SLAP_CAUCR
5	237.5	5.0	881	1	YJH8_YEAST
6	234	4.9	1645	1	OMPB_RICPR
7	232	4.9	1643	1	YIO9_YEAST
8	230	4.8	995	1	YFAL_ECOLI
9	230	4.8	1250	1	YFAL_ECOLI
10	226.5	4.7	1609	1	YFAL_ECOLI
11	224.5	4.7	2003	1	YDBA_ECOLI
12	223.5	4.7	1656	1	OMPB_RICUA
13	222	4.6	1041	1	EGT2_YEAST
14	220.5	4.6	725	1	AGAI_YEAST
15	218.5	4.6	1140	1	YMG6_YEAST
16	215	4.5	1848	1	CBPA_CLOCL
17	214	4.5	1569	1	YRJA_ECOLI
18	210.5	4.4	1120	1	STRF_ECOLI
19	210	4.4	1654	1	OMPB_RICRI
20	209	4.4	1039	1	AG43_ECOLI
21	208.5	4.4	1300	1	120K_RICRI
22	206	4.3	1306	1	MSB2_YEAST
23	205	4.3	1385	1	AIDN_ECOLI
24	205	4.3	1419	1	ALAI_CANAL
25	204.5	4.3	1150	1	APWQ_PIG
26	200.5	4.2	749	1	TROP_HUMAN
27	200.5	4.2	1169	1	YK82_YEAST
28	199.5	4.2	933	1	SLAP_CAMEF
29	197.5	4.1	796	1	YS8A_CAMEL
30	192	4.0	1577	1	HLVA_PROMI
31	191.5	4.0	959	1	N100_YEAST
32	189.5	4.0	610	1	MUC4_HUMAN
33	189	4.0	1251	1	YOUS_CAMEL

34	186	3.9	797	1	VGIX_HSVB	P28968 equine hep
35	186	3.9	1161	1	YJ9P_YEAST	P47179 saccharomyc
36	185	3.9	598	1	YAEF_SCHPO	Q10168 schizosach
37	185	3.9	1322	1	ICPA_PANAN	P20469 pantoea ana
38	185	3.9	1367	1	AMTH_YEAST	P08640 saccharomyc
39	183.5	3.8	1409	1	HAP1_HAELIN	P44596 haemophilus
40	182.5	3.8	774	1	STF_LAMB	P03764 bacterioph
41	182.5	3.8	3591	1	FHAB_BORPE	P12255 boideella
42	180.5	3.8	2090	1	N214_HUMAN	P35658 homo sapien
43	180.5	3.8	2329	1	YS89_CAMEL	O09624 caenorhabd
44	179.5	3.8	532	1	SP67_DICDI	P22698 dictyostei
45	179.5	3.8	1325	1	YDEK_ECOLI	P32051 escherichia

ALIGNMENTS

RESULT	1	ALIGNMENTS
OMPE_CHLTR	STANDARD:	PRT: 964 AA.
ID	OMPE_CHLTR	
AC	O84877;	
DT	30-MAY-2000 (Rel. 39, Created)	
DI	30-MAY-2000 (Rel. 39, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	POTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.	
GN	PMPE OR CT869.	
OS	Chlamydia trachomatis.	
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=D/JW-3/CX;	
RX	MEDLINE: 99000809;	
RA	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,	
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,	
RA	Davis R.W.;	
RT	"Genome sequence of an obligate intracellular pathogen of humans:	
RT	Chlamydia trachomatis."	
RL	Science 282:754-759(1998).	
CC	-1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)	
CC	(POTENTIAL).	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
DR	EMBL: AE001360; AAC68467.1; -	
KW	Outer membrane; Signal.	
FT	SIGNAL	
FT	CHAIN	POTATIVE OUTER MEMBRANE PROTEIN E.
FT	SEQUENCE	964 AA; 104703 MW; 1B998A7D2E571CE2 CRC64;
QY	Query Match	11.3%; Score 541.5; DB 1: Length 964;
QY	Best Local Similarity	25.5%; Pred. No. 7,6e-23;
QY	Matches 245; Conservative 138; Mismatches 400; Indels 177; Gaps 47;	
QY	62 LENIPOTGAIT-----KSFNNKCGDLFTFGNGNSLFTQVAGTVAGAAVSSVVDKST 117	
QY	68 LKTPBEGAAVITTDLSFDDKEIYRAKN-----LPPESGAGVSPNSPTVEIND 122	
QY	118 TF--IIFSS-----LSFIASP--GSSITTGKAVSCSTSLTKNVSL--FSKNFSTON 167	
QY	123 TILPVIFENNCTCRLETPWNPYADKIREG-GAIHQ--NLXINHHADVGEKFNFSYQ 179	
QY	168 GGA1-TAKTLSLGTMSLFSFN-----TSKKGAIGTSDALITTGNGEVSFSDNTS 221	
QY	180 GGA1STANTFVVSNSOCLFMDNICIQTNTAGKGAITAGSNGSFSNNCDLFFINNA 239	
QY	222 SDGGAIFTEASVTISNNAKVSFIDNK-----VTGASSSTTGDMSGAIGAYKTSTDTKYT 277	

```
Db 240 C-AGGAIFSPICSLGRGNIVFYNNRCKNVEETASSEAS---DGCAL-----KVTTRLD 290
Oy 278 LTGNQ-MLFFSNNTSTTAGCAIYVKKLEIASGGLTFPSNSVNGCTAPGCAIADDSGE 336
Db 291 YVGNRGRIFFSNIRKNYGAIAYPVLTVDNGPTFFINNINN---KGGAIYIDGTSN 346
Oy 337 LLSADSGDIVLGNVTSTT--PGTN-----RSSIDLGSAKMTALRSAGRAIYF 387
Db 347 SKTISADRAHIINENIYVITNANGTSTANPPRRNATVASSGAILLAGASSQMLIYF 406
Oy 388 DPTTGSSTTVTVDLKVNTPADSALQYTGNIIFTEGKLESTEADSKMLTSLQPLVL 447
Db 407 DPEVSNAGVSEYFNK-----EADQ-----TGSVFCATVNSAD-FHQRLQKTPAPLTL 457
Oy 448 SGGTSLKRGVLTQATQOAD-----SLEMGVGTLEPADSTINNLYNIS 497
Db 458 SNGFLICIEDHAQLTYNRFQOTGGVSLGNGAVLSCYKNGTGSASNA-SITLKHIGLNS 516
Oy 498 SI--DGAK--KAKIETKATSKMLT--LSGTTILDPDTGFYENHSILRNPQSDILELKAS 551
Db 517 SILKSGAEIPLWVEPTNSNNYTADTATFSLDKLSLIDYGSPESTDLTHALSS 576
Oy 552 GTVTSTAVTPD-PINGEKF-----HYGQGTWGPVWG-----TGASTATAT-- 591
Db 577 QPMLSTSEASDNLQLOSENIDFSGLVNPHYGMQGLW---TWGNAKTQDPPASSATITDPQ 633
Oy 592 -----FWMTKGYIPNPERIGSLVPSLW-NAFIDISLHMLTANEGLOGDNA 640
Db 634 KNRHRHRLTLTLWPAGYVSPKHSPLANTLWGMMLATSLKNSALTLPSG---HP 689
Oy 641 FW---CAGLSNFFHKDSTKTRGRFRLSGGY---VIGNLHTCSPKILSAFCQLFGDR 694
Db 690 FNGITGGGLGMVYDDPRNRHHPGFHRSSGYSAGMTAGTHTFSLK-----ISQITTKLN 744
Oy 695 DYFAKNOCTVYGGILYOHNETYISLPCKLRPCSIYPTIPIVLF--GNLSTYTHD 751
Db 745 ERY-ARKN--VSSKNYSCQGEMLFSL-----OEGFLTLVGLYSYGDNCHMFYRQ 793
Oy 752 NDLKRYTTY--PTYKSGNDSPALFEGGRAPICIDESALREQVWPEMKLOFYVAHOG 809
Db 794 GENLTSOGTFRSQTWGGAFFDLPMKPRGSTHILT-----APFLGALGTSSLSH 843
Oy 810 FKEQGTAREFGS-SRLVNLALPIGRPKESDCODAT-----YNLLGTVDLYRSND 863
Db 844 FTEVGAYPRSFSTKPTLNLVLPVIGK---GSEMMATHRPQAWYELAYQVLYRQEG 899
Oy 864 CTTTIRISGDSKRTGTNLARQALYLRAGN-----HFCFNSNFPAESQSFELRG 913
Db 900 IAAOLASKGIWFGSGSPSRHAMSXYISQOTQPLSWTLHFQYH-GFYSSSTFCNYLNG 958

RESULT 2
OMPF_CHLTR STANDARD: PRT: 1034 AA.
AC P38008; 084878;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN P PRECURSOR.
GN PMFP OR CT870.
OC Chlamydia trachomatis.
OS Chlamydia: Chlamydiales; Chlamydiales; Chlamydia.
RN [1]
RC STRAIN=D/W-3/CX.
RX MEDLINE; 99000809.
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RL Chlamydia trachomatis."
Science 282:754-759(1998).
```

```
RN [2]
RC SEQUENCE OF 26-35.
RP STRAIN=L2/434/BU.
RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christlanen G., Birkelund S., Virelou E., Ratti G.,
RA Pallini V.:
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
CC -i SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC
CC This SWISS PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE001360; AAC68468.1; -.
KM Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR->FH (IN REF. 2)
SQ SEQUENCE 1034 AA; 112392 MW; 445FF4C3D463AE7 CRC64;
```

Query Match 10.3%; Score 490; DB 1; Length 1034;
Best local similarity 23.5%; Pred. No. 6,1e-20;
Matches 258; Conservative 143; Mismatches 428; Indels 268; Gaps 49;

```
Oy 5 PPKVEFS---TFAIPLSMIAETETVIDSSAFDGNKGNFSVRESQDAGTYLTKGNV 60
Db 33 FRRFFFSREIOFVDPDPAFLTAQNIIVLS--NLSNGTGACTIONTQ--TQIFSNV 86
Oy 61 TLENIPGCTATKCEKFNNTKGDLETFNGNSLLEFQYDAGTVAGAAVSSVVDKSTFEI 120
Db 87 NT-----TADSGAFPMVTTSTFASDNANLFCNNCTYNNKGGAIRSG--GPIFL 136
Oy 121 GFSLSF---IASPSSITTGK-----GAVSGTGLSTLTKVNSLFSKNFSTDNGA 170
Db 137 NNQDPLFNNISAGKAYVGTGDHNEKNKGALYAT--ITLGNRTLAFINMSDCCGA 194
Oy 171 ITAKT-LSLTGTMALFSEN-----TSSKKGAI-QTSDALTITGNGEVSFS 217
Db 195 ISADQIISITTVKILFLENHHTLNHPYTOAENARAGACSRDLCSINNSGPVIFN 254
Oy 218 DNTSSDGAATFTEASVLTISNAKVSFIDNKVTGASSSTGDMSGCAICAKYTSRDTVT 277
Db 255 YN-QCGKGAISATRCVIDNKKERTIFSNSSLGWSOSSAS--NGAI-----QTGQFT 307
Oy 278 LTGNQ-MLFFSNNTSTTAGCAIYVKKLEIASGGLTFPSNSVNGCTA-----PKGAIAT 331
Db 308 LRNNKGIYFDNSTATHAGAINCGYIDIRDNGPYFLNNSAAMGAARNLKPR----- 361
Oy 332 EDGSELASADSGDIVLGNVTSTTBT-----NNSIDLGSAKMTALRSAG 381
Db 362 --SATNYIHTGTGDIV--NNWVFILDNLLGKRKLFHINNETHPTPY---LSLAKKD 414
Oy 382 RAIYFYD-----PITGSSSTVTVDLKVN-ETPADSALOYTGNIIFTEGKL 426
Db 415 TRIFYDLPQERKVENTSNPPSPSTSRNITT---VNPET-----EFGAVVFSTYQM 464
Oy 427 SETEADSKNLTSK-----LLQPYTLSGTSLKRGVTLQ--TQAFQOADSRLMDVGT 479
Db 465 S-----SDIRTLMGKEHNNIKRAPTLTKRGTLAIDDALELFNPFONPRLSLALSGA 520
Oy 480 TLEPADISTIN---NLVINISSIDGAKKAKIE-----TATSKNLTLGTTLLDPTG 529
Db 521 TLTVGKHKMLITNLGVLPIILKEGKSPCIRVNPQDMQNTGTGQTPSSSTSSISTPMI 580
Oy 530 TFEYENHSL--RNPOS-YQILEL---KASGIVTSTAVVPPDPMGEKF-----HY 571
Db 581 IFNGRLSTVDENYESVDSDMLSRKAQQLILSTETITDQGLDSNWSQSLNTSLSPHY 640
```

OY 572 GYOSTMGPIW-----582
 DB 641 GYOGMLTP-WNITTTTITLNNSSAPTSATSIAEOKTSETFPSTNTTASIPNIKASA 699
 OY 583 -GTGAST-----TATFNMKTGYIPNPERIGSLVNSL---WNAFIDISSIHY 626
 DB 700 GSGGASANSSEVITKHTLVNNAAPGYIYDPRRGDILNLSLVHSGRNMWMLRSL-- 757
 OY 627 LMETANGLGDRAPWCAGLSNFFHKDSTKRGRFRLSGGYI---GGLNLTGSDKI 681
 DB 758 LPDMSMELG-----AATLTFTKOOKRLSYHGYSSASKGYTSSOASGAHGH----- 805
 OY 682 LSAFCQLFGRDRPYFAKN-----OGTVYGGTLYOHNEHYISLPKLRCSISYVPT 736
 DB 806 ---KFLSFSSOSSDKMEKETNNRLSSRYISALCFEH-----PMF 843
 OY 737 IYVLFSGNLSYTHNDLTKYTYTPYKSGWNSDFALFEGRAPIC-LDESALFEOYM 795
 DB 844 DRILIGAACNVCYTHMNRSPYTKSKGFHSTTL---GASLNCGLRDSNPLRSIM 898
 OY 796 --PPMKLOFYVAHOEGKEGTEAREFGSSRL-VNALPIGRFDSDC-QDATYMLTL 851
 DB 899 LTPPAQLFRTEPASJRESDDLRLFTLEQAHRAVSPGICKAYSSDTPWTLSEWEL 958
 OY 852 GYTVDLRSPDCTTTLRISGDSWKTFTNLARQALVLRAGNFCFNSNEAFSGSFEL 911
 DB 959 AYOPLTWKRRLNTLTLIONNGSWTNTPLAKHSFYGR-GSHSLKFSHLKLFANYOAEV 1017
 OY 912 RGSRRNVLDGAKYQF 928
 DB 1018 ATSTVSHYNAGALVF 1034

RESULT 3
 190K_RICRI STANDARD; PRT: 2249 AA.

AC P15921;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R.
 RX MEDLINE: 90354033.
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.J.,
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly
 RT repeated, near-identical sequences.";
 RL Infect. Immun. 58:2760-2769(1990).
 CC -FUNCTION: ELICITS PROTECTIVE IMMUNITY.
 CC -DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
 CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
 CC -This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 DR EMBL: M31227; AAA26380.1; -.
 DR PIR: A41477; A41477.
 KW Antigen; Repeat; Signal; S-layer; Glycoprotein.
 FT SIGNAL 1 28
 FT CHAIN 29 2249 POTENTIAL.
 FT DOMAIN 212 1180 190 KDA ANTIGEN.
 FT REPEAT 212 286 13 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 287 358 A (TYPE I).
 FT REPEAT B (TYPE II).

FT REPEAT 359 430 C (TYPE II).
 FT REPEAT 431 505 D (TYPE I).
 FT REPEAT 506 577 E (TYPE II).
 FT REPEAT 578 652 F (TYPE I).
 FT REPEAT 653 724 G (TYPE II).
 FT REPEAT 725 799 H (TYPE I).
 FT REPEAT 800 874 I (TYPE I).
 FT REPEAT 875 949 J (TYPE I).
 FT REPEAT 950 1021 K (TYPE II).
 FT REPEAT 1022 1093 L (TYPE II).
 FT REPEAT 1094 1165 M (TYPE II).
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
 SQ SEQUENCE 2249 AA; 224333 MM; A9D6646C089DF087 CRC64.

Query Match 5.6%; Score 269; DB 1; Length 2249;
 Best Local Similarity 22.9%; Pred. No. 3.4e-07;
 Matches 203; Conservative 100; Mismatches 324; Indels 258; Gaps 43;

OY 60 VTLENIPGCTAITKSCFNNTKGDITFTGN-----GNSLFFQYDAGTVAGAAVNSSV 112
 DB 396 VTFTN-PVVVTGAIDNTGNANNNGIVFTGNTVTDIGNTNALATVNVG--AGTATLGA 452
 OY 113 VDKSTT-----FIGSSLSFIASPGSSITTG-----KGAVSCSGSLSTKNV 155
 DB 453 VIKATTKLTNAASVLTLTNANAVLTGALDNTTGGDNVGVNLNGLASQYTGIGINTSL 512
 OY 156 SLF-FSKNFSTDNGALITAKTLSTGTMSALFS-----ENTSCKKGAIOTSDAL 205
 DB 513 ATISVAGATTLGGAIVKATTKLTDAASAVKFTNPVVVTGALDNTGNANNNGIVFTGNS 572
 OY 206 TITNGOGE-----VSSDNTSSDGAAT-----FTEA-----SVTISNAKSFIDN-- 247
 DB 573 TVTGDIGNTNSLATISVAGATLGGAVIKATTKLTNAASVLTLTNANAVLTGALDNTT 632
 OY 248 -----KVTGASSSTTGDM-----GGAICAVKTSDFKVTILGNQ 282
 DB 633 GGDNVGVNLNGLASQYTGIDIGNTNSLATISVAGATLGGAVI---KATTKKIT-NAVS 688
 OY 283 MLFNSNTSTTACGAIYKKLELASGGLLFSRNS-----VNGTAPKG 326
 DB 689 AVKFTNPVVVT--GAI--DSTGNANNNGIVFTGNTVTDIGNTNALATVNVGATLIG 744
 OY 327 GAIAIEDSGELS-----LSADSGDIVLGNVTSTPGTRRSI-----DLGTS 370
 DB 745 GAVIKATTKLTNAASVLTLTNANAVLTG- AIDNTTGGDNVGVNLNGLASQYTGIDIGN 803
 OY 371 AKMTALRSAGRAIYFVDPITTGSSSTTVTDVLRVNETPADSALOYT-GNIIFTGKLESET 429
 DB 804 NSLATISVAGTA-----TLGAVIKATTKLT--TNAASVLTLTNANAVLTG-AVDNT 853
 OY 430 EAADSKRLTSLKLPVTLTSGTSLKHGVLQTOATQOADSLEMDVGTTEPATDSTI 489
 DB 854 TGGDNV-----GVNLN-NGALSQYTGIDIGNTNSLATISVG-----AGTATL 893
 OY 490 NNIVINISIDGAKKAIETKATSKNLTLSGT---TLDPTGTFFENHSLR-----N 539
 DB 894 GGAVIKATTKLTNAASVLT-LTNANAVLTGALDNTTGGDNVGVNLNGLASQYTGIDIGN 952
 OY 540 POSYDILELKASGTVT-----STAVTPDPIMERKFNHYOGTGWPIWGTGA----- 586
 DB 953 TNSLATISVGA-GTATLGGAVIKATTKLTLDASAVKF-----TNPVV-VTGALDNTG 1003
 OY 587 -----STAFENMTKGYIPNPERI-----GSLVNSLMNAFIDISSIHYLM 628
 DB 1004 NANNNGIVFTGNTVGNVGNLTALVNVGAGLLOVGGVAAANINLTLDNASAVFTN 1063
 OY 629 ETANEGLOGDRAPWCAGLSN---FFHKDSTKTRGRFRLSGGYVIGNLHTCSIDLISA 684
 DB 1064 PVVVTG-----AIDNTGNANNNGIVFTGNSVTV-----GNV--GNTNALATVNVGA 1107
 OY 685 APCQLGRDRDYPVAAKNOCGYIGTLYOHNEHYISLPCKLRCSISYVPTLPVLPFSGN 744

DB 1108 GLOVOG-----GVKANTINLTNDN-----ASAVFTNPVVVTGA 1142

QY 745 LSVTHDNDLTKRTYTPYKSGWGNDS-----FALEGG 779

DB 1143 IDNTGNANGIVTFTGNSTVTGDTGNTMALATVYVAGAGITLLOAGG 1187

RESULT 4

SLAP_CAUCR STANDARD; PRT: 1025 AA.

AC P35828: Q46015; (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).

GN R5AA.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21: 646-660 AND 1021-1025.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE: 93007489.

RA Gilchrist A., Fisher J.A., Smit J.K.;

RT "Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus paracrystalline surface layer protein.";

RT Can. J. Microbiol. 38:193-202(1992).

RN [2]

RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE: 89008089.

RA Fisher J.A., Smit J.K., Agabian N.;

RT "Transcriptional analysis of the major surface array gene of Caulobacter crescentus.";

RT J. Bacteriol. 170:4706-4713(1988).

RN [3]

RP CHARACTERIZATION.

RC STRAIN-CB15;

RX MEDLINE: 98292737.

RA Awram P., Smit J.K.;

RT "The Caulobacter crescentus paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus.";

RT J. Bacteriol. 180:3062-3069(1998).

CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.

CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF062345; AAC38665.1; -

DR HSSP: F01549; 2MCM.

DR INTERPRO: IPR001343; -

DR PFAM: PF00353; hemolysinCabinid; 1.

DR PRINTS: PR00313; CABNDNGRPT.

KW Cell wall; S-layer; Calcium-binding.

FT INIT_MEN 0

FT SEQUENCE 1025 AA; 98209 MW; AFC8B519820B1A5F CRC64;

Query Match 5.1%; Score 244.5; DB 1: Length 1025;

Best Local Similarity 24.2%; Pred. No. 2,6e-06;

Matches 179; Conservative 86; Mismatches 257; Indels 217; Gaps 33;

QY 12 TEAIFPLSMIAETVLVDSASFDGNKGNFSVRESQEDAGTYYLEKGNVLTENIP----G 67

DB 268 FEVAEEVAGAAATLVGDFLLSGAGT-----DVLNVAQAAYALPTG-VTISGIETMNT 321

QY 68 TGTATITKSCFNNTRKDLFFFTGNGNSLFFOTYDAG-----TVGAANSSVVDKSTFI 120

DB 322 SGAATTLNTSSGVTS-LTALNTTSGAAQYTAGAGNLTAATAQAANNVAVD----- 374

QY 121 GFSSLFIASPGSSITTKGAVSCSTGSLTKNVLFSKNFSTDNCGATYAKTSLTG 180

DB 375 GRANTVASTGVTSGETTVGANSASGVTSV-----SVANSSTTTGAIATG 422

QY 181 TTMALFSENTSSKKGAIGTSDALITGNOG--EVSESDNTSDSGAIF--TEASVTI 236

DB 423 GL--AVTAAQTGMAVNNTLLQADVTVGNSSTAVTVTQTAATAATAGATVAGRVGAVTI 480

QY 237 SNNA-----KVSFIDNKVTGA-----SSSTTDMSG-----GALCAVYTSYDT 274

DB 481 TDSAAASATTACKIATVVLGSGAATIDTSALTIVNLGSGTSLGIGGALTA----- 533

QY 275 KYTLGNOMLFSNNTSTTAGAGIYVKKLELASGGLTFSRNSVNGTAPKG-GAIAIED 333

DB 534 --TPTANTLTLLVNGLTIT--CAITDSEAADG-----FTTININGSTASSSTIASLVAD 585

QY 334 SGEISLSDSGDIV-----FLGNTVTS-----TPQTNRSSIDLGTSA 371

DB 586 ATTILNSGARVYTISSHAAALTGITVNSGATLGAEIANGLVFTGAGRDSILLGAT 645

QY 372 KMTALRSAGRAIYFYDDPTTGSSTTV-----TVILKVN----- 405

DB 646 K--AIVMGAG-----DDTVYSSATLGGSGVNGGDDGVLANVNGSSFSADPAFGCF 697

QY 406 ET--PADSALOYTGNI--FTGKLSSETAA----- 432

DB 698 ETLRAAGAAAGSHNANGFTALQIGATGATTTVVAVNVGLVLAATGTTVTLANAT 757

QY 433 ---DSKNITSLKLPVTLTSGTSL-----KHGVTLQTA-----FT 466

DB 758 GTSDVENLT--LSSSALAAGTVALAGVETVNIATDNTTAHVDTLLQATSAKSIYVT 815

QY 467 QQADSRLEMDVITTEPPADTSI-----NVLVINSISDQAKKAKIEKAKSK 514

DB 816 GNAGINLTNGTAVTSPDASAVTGAIVFVSANTTGEVVTITRGAGADSLTGSATA 875

QY 515 NLTL---SGTILLDPGTGFYFNHSLRNPQSYDIELKASGVSTAVTAPPIMGEKHY 571

DB 876 NPTIIGAGADTLVYTG--DTFTGTGADIPDINAIGSTAFVTTTDAVADKDL 931

QY 572 GYGTGWPIWGT-GASTT 589

DB 932 VGISTNGAIDAGAFGAATV 950

RESULT 5

ID YJH8_YEAST STANDARD; PRT: 881 AA.

AC P47033;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCPL60-SMC3 INTERGENIC REGION.

GN YJL078C OR J1027.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C;

RX MEDLINE: 96093911.

RA Miosga T., Schaaff-Gerstenschlaeger I., Chaltatzis N., Baur A.,

RA Botes E., Fournier C., Schmitt S., Velden C., Wilhelm N.,

RA Zimmermann F.K.;

"Sequence analysis of a 33.1 kb fragment from the left arm of

RT Saccharomyces cerevisiae chromosome X, including putative proteins
 RT with leucine zipper, a fungal Zn(II)-2-Cys6 binuclear cluster domain
 RL and a putative alpha 2-SCB-alpha 2 binding site."
 CC yeast 11:681-689(1995).
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5: FUNGI SC7/SC14 AND PLANTS PR-1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sid.ch).
 CC
 CC EMBL: 249353; CAA69370.1; -
 CC EMBL: X83502; CAA58492.1; -
 CC EMBL: X88851; CAA51314.1; -
 CC HSSP: P04284; 1CFE.
 CC SGD: S0003614; YJ0078C.
 CC INTERPRO: IPR001283; -
 CC PFAM: PF00188; SCP_1.
 CC PRINTS: PR00837; V5TPYLK.
 CC PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 CC PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 CC Hypothetical protein
 FT DOMAIN 166 379 ALA/SER/THR-RICH.
 FT SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match 5.0%; Score 237.5; DB 1; Length 881;
 Best Local Similarity 23.6%; Pred. No. 5.2e-06;
 Matches 127; Conservative 86; Mismatches 209; Indels 117; Gaps 22;

39 GNFS--VRESQEDAGTYLFGKN-----VLENIIPG-----TGAITK 74
 112 GHPTQVWKKSTAEICGCKYKCGTWMNYIVCSYVNPGRNGLGEFAEEVPELISTVSSSS 171
 75 SCNNNTKGLDTFTGNGNSLLFOTVDAG--TVGAANVSVVVKSTTTFIFSSLSFTASP 131
 172 SSSSTSTSDVST--ISSSTMPAVAOGYTTVSSAASSSL--KSTTINDAKATITLASS 227
 132 GSSITTGKAVSGSTGSLSLTKNVSLLFSKFNSTDNCAITAKTLSEGTMSALFSEMT 191
 228 STYTSTSTESVGSYVSSASSSV-----TTSYATSSSTVSSDATTSTTSSVATSSST 283
 192 SSKKGAIGTSDALTTTNGOGEVFSFSDNTSSDSGAALFTEASVTISNNAKVSFIDKVTG 251
 284 TS-----SDP--TSTTAASSSDPSSAASSASTENAASSSAISSSSSVSA 332
 252 ASSSTTGDSMGCAICAVKSTDTKVTLLTGNQ--LFSNNTSTTAGAIVYKLELASG 309
 333 PLST--LTSTASSSVTSVNSVSKFAN----- 361
 310 LTFSSNSVNGTAPKGAIAIED--SGELSLSDSGDIVLGNTVTSTPGTNR--SID 366
 362 -TFESAQTTSVSAISSVAAADIDGISTKEATSS--VSEHTSVTATNAAOYATR 416
 367 LGTSAMT--ALBSAAGRAIIFYDPTTGSGTVDYLVAKNEFPADALQYTNITFTGE 424
 417 LGSSSSSSSAVSSAVSOSVLSVAVNVDVSTVSSSTAHTKDTA--TTSVASES 473
 425 KLESTEAADS--KNLTSKLQPTVLSGGTSLK--HGVTLQTO--AFYQOQDSLENDV-- 477
 474 ITSETQASSTKKNISNKAATSSSIYSNSASVGHGYTAAEYAITSESSALATSVPA 533
 478 -----GTLLEPADSTINNLVINISSIDAKKAKIETKATSKNLTLSGTTLLDPT 528
 534 TNCSSIVKTTLENSSTYTTAIT-----KSTTLTATANNSTRAATAVTITDPT 582

RESULT 6
 OMPB_RICTY

ID OMPB_RICTY STANDARD: PRT: 1645 AA.
 AC P96989;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN SLP.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-WILMINGTON;
 RX MEDLINE: 94040787.
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene", encoding the crystalline
 RT surface layer protein of Rickettsia typhi.";
 RL Gene 133:129-133(1993).
 RN (2)
 RP PARTIAL SEQUENCE.
 RC STRAIN-WILMINGTON;
 RX MEDLINE: 92114896.
 RA Ching W.-M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNB fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN (3)
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE: 92104668.
 RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent
 RT mutant deficient in processing.";
 RL Infect. Immun. 60:158-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sid.ch/announce/>
 CC or send an email to license@isb-sid.ch).
 CC
 CC EMBL: L04661; AAB48987.1; -
 CC Antigen: S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 FT SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 4.9%; Score 234; DB 1; Length 1645;
 Best Local Similarity 21.6%; Pred. No. 1.9e-05;
 Matches 206; Conservative 129; Mismatches 356; Indels 264; Gaps 50;

20 MATENVLDSASFQCNKNG-----NFSVRESQEDAGTYLFGK-NVYLENIPG 69
 678 IVAADPINTDTALADGTNLGSAESPLSIHFAIKRANGDS-ILHIGKGVNLYANNITTTD 736

FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA: 169854 MW: 735FDF392E6346CC CRC64:

Query Match 4.9%; Score 232; DB 1; Length 1643;
 Best Local Similarity 20.9%; Pred. No. 2.5e-05;

Matches 183; Conservative 130; Mismatches 330; Indels 232; Gaps 43;

23 TETVLDSASFQDGNKNGNFVSRESQEDAGTTLFKGNVT-----
 748 TSVISGTVGGQGHKLNNLL-----DNGTIVKRLGDTTNGGKIKGKSLIQSNVYT 802
 62 --LENIPTGT-----AATKSCFNNTKGDLT--GN----- 89
 803 DHEASADNTLLEFVNDPTITVLNKGAGFVGLKOVIIISGPGIVNEIGNWIGVIA 862
 90 GNSLLEFQTVAG-----TVAGAAVNSVVDKSTTFIFGSSLSFIASPSSEI 135
 863 ANSISFENASLGTSFLPSGTPLDVLTKSTVGCTVNDNNAPIVAVVSGIDSMINQOI 922
 136 TTGKAGVSCGTG--LSLTKNVSLFSKNFSTDNCGAITANTLSLTG-----TTMSALFS 188
 923 GDRKNIITALLSGDINSITVANTLYSGIRTKNNQG-----TVLSGGMPPNPGIYGL 978
 189 ENTSKKGAIOTSDALITINGOG-----EVSFSDNTSSDGAIAFT--EASVT---ISN 238
 979 ENSPKIKOVYTFITD-----YNNIGSIANNVTINDVYTLTGAGTDPKATITLGSVNG 1034
 239 NAKVSFIDNKVTG-----ASSSTTGDMG--GGAICAKYKSTDT--KVLTGNOMLFS 287
 1035 NAWRFVDSFSDPRSMIVATQANKGTVTLGNALVNSISLDIPVASVFTG----- 1087
 288 NNTSTAGGAIIYKKLELAGGILFSRNSVNGSTAPKGAIAIEDSGELSDSCDIY 347
 1088 NDSGAGLOGNIYSONIDFGYINLTILNSNVLG-----GGTFAI--NGEIDL--TNLLI 1138
 348 FLGNVTSTPGTNRSSIDLTGTSAKMTALRSAGR--AIIFYDPTTGSSTVTDVKVN 405
 1139 FANGT---STWGDNTSI-----STTLNVSSNIGOVVIAEAOVNAITTTGTTIKIO 1187
 406 E-----TPADSAIQ-----YTGNITFTGKLESEAD--SKNLTSKL 441
 1188 DNANANESGTOAYTLIOGARGFNGTLCAPNFAVYTSNIFVAKYELLIRDSNODYLTINDV 1247
 442 LQVTLISGGLSKH--GVYLOTOAFQOADSRLKMDVGTLLERADSTINNVI----- 494
 1248 LNVYTTAVGSAIANAGVS-----QNISRCLESTINAAVNNMLANDPS 1292
 495 NISSIDGAKKAKIETKATSKNL--TLISGTYLLDPTGTFEYENSLRNFOSYDILELKASG 552
 1293 DVATFAGAIATDISAANTYTNLNDQTOPLLSNRGLT--KLSNAETSDVAG--SATG 1348
 553 TVTSTAVTPPIGKEFHGYOQTWGPYWGTA---TAFPMNTKTYIYIPNERIGS 608
 1349 AVSS-----GDEAEVSY--GVMAKPFYNAIEODKKGAGIAGKAKTGVVVGDLTLAS 1398
 609 LVNPSLWNAFIDISLHYLMEFANEGLOGDRATFACAGLSNFHKDKTKTRGRFRHISGCV 668
 1399 --DNLMGAIGITKTDIKHODYKKGDKT---INGLSLSLY-----GSGQLKNF 1444
 669 VIGGNLHTCSDKILSAA--FCOLFRDRDYFAAKN--QGTVYGGTLYYOHNETYISLPCK 724
 1445 FAAGNAIFTLNKKVKSQRVFEFSGNKMKSQIAAGNYDNMTFGGNLIFGVD--YNANPNV 1502
 725 L--RPGSLSVPTPEIPLVLESGNLSYTHDNDLTK 757
 1503 LVTPMAGLSY-----LKSSENENYKETGTTVANK 1530

RESULT 8
 Y109_YEAST STANDARD: PRT: 995 AA.

P40442;
 01-FEB-1995 (Rel. 31, Created)
 01-FEB-1995 (Rel. 31, Last sequence update)
 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHEICAL 99.7 KDA PROTEIN IN SDLI 5'REGION PRECURSOR.
 GN YIL169C OR Y19402.07C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 Raulier C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 Gentles S., Hamlyn N., Horenell T.S., Hunt S., Jagsels K., Jones M.,
 RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 246921; CAA87023.1; -
 DR SGD; S0001431; YIL169C.
 KM Hypothetical protein: Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 995
 FT CARBOHYD 26 28
 FT CARBOHYD 35 35
 FT CARBOHYD 468 468
 FT CARBOHYD 664 664
 FT SEQUENCE 995 AA: 99735 MW: F63E287A03F137EC CRC64;

Query Match 4.8%; Score 230; DB 1; Length 995;
 Best Local Similarity 21.2%; Pred. No. 1.6e-05;
 Matches 180; Conservative 136; Mismatches 355; Indels 180; Gaps 35;

11 STFAIFPLMIATETVLDSASFQDGNKNGNFVSRESQEDAGTTLFKGNVTLENIPTGT 70
 10 AALALVYSOALGOYYSNSTISSNSSFVSQSS-----SGSVSI---SSSI 55
 71 AATKSCFNNTKGDLTFTGNGNSLLEFQTVAGVAGAANSVVDKSTTFIFGSSLSFIAS 130
 56 AETSSATDILSITQASASTSGVSSV--GPRSSSVSSVSQSS-----SSVSQSS 107
 131 PGSSITTGKAGVSCGTG--LSLTKNVSLFSKNFSTDNCGAITAKTLTLGTTMALFSE 189
 108 SVSOSSSASDVSSVSQASSTSDVSSVSQSSASDVSSVSQSSASDVSSVSQSS 167
 190 NTSKKGAIOTSDALITIGN-----QGEVSFSDNTSSDGAIAFTFEASV 234
 168 SASASADVSSVSQASSTSDVSSVSQSSASDVSSVSQSSASDVSSVSQSSAS 227
 235 TISNNAKVSFIDNKVTGASSSTTGDMG--GGAICAYKTSITDKVTLTG----- 280
 228 TSDVSSVSQASSTSGVSSGOSVSSASGSSSPQSTASASTGASATNSLSITS 287
 281 --NOMLFSNNSTTAGGAIIYKKLELAGGILFSRNSVNGSTAPKGA--IAIEDSGEL 337
 288 SASASATASNSLSISSDGTIYLPPTTI--SGDLTLGKVATEGVVAAAKTLLDGKRY 346
 338 SLSDA---SGDIYFLGNTVSTTPTG--NRSSIDLTGSAKMTALRSAGRA--IYFDPT 391
 347 SFSADLKVYGD--LVKKRKEIYPTGEFIDISGENPDVTONFNAEBAASATISTPTSS 404
 392 TGSSTVTVDVL-----KVNETPADSALOYT--GNIITGKLESEPADSK----- 435

[illegible]


```

OY 265 ICAYTSIDTK-----VTLTGNOMLFEFNNSTSTGAGAIYVKKLEL-----AS 307
Db 208 IDVTDNNSGDSKHPSGTIVNN--TAFNTNAGYGAITNTSYPALIYDISVDSYSON 265
OY 308 GGLTIFSRNSVNG---GTA PKGGAIAIEDSGELSLSDSGDIVLGNTVSTTPGTNRSS 364
Db 266 GGVLVDENNNSAAGYDGPSSAAGCFMYLGLSEVTEDTADGKTLVIGNT-----ENDGA 318
OY 365 IDLGRSAKMTALRSAGRAIIFYDPDITTGSGSTTVYDVLVKNETPDSAALOYGNILFTGE 424
Db 319 VD---SIAGTGILITGSG---DLVLNANDNDFTGEMOJEN--GEVTLGRNSLNVND 369
OY 425 KLSETEADSKMLTSEKLLDPVTLISGGLTSLKAGVLQT--QAFTOADSRLMEMDYTLE 482
Db 370 THCQDDPDQCYGLT---IGSIDQYNQALNANSGTQGTIVHMLTGRONTLWDAG----422
OY 483 PADTITNNLVNIS----IDGAKKAKETK-----ATSKULLTSGITLLDPTGTFFE 533
Db 423 -----GNVTYNQSSFAGIIEGAGCLTIAONGSYVLAAQNSMALTLGIVDDGAVLSLE 475
OY 534 NNSL-----RWPOGY-----DLELK-ASGVTSTAVTPDDI 564
Db 476 GDAALTLALQDDPQIVLNGVGLDLSDFSTWOSGTSYNDGLEVSOGSGTIVISQDVYDLA 535
OY 565 MGEKPHYGYOGTGRPIY 581
Db 536 GGDNLHIGDGDKGYV 552

RESULT 10
FIG2_YEAST
ID FIG2_YEAST STANDARD: PRT: 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE FACTOR INDUCED GENE 2.
GN FIG2 OR YCR089W OR YCR089W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
RN Saccharomycetaceae; Saccharomyces.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92397594.
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RT cerevisiae chromosome III contains two new open reading frames.";
RL Yeast 8:569-575(1992).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -!- INDUCTION: BY MATING PHEROMONES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59720; CAA42254.1; -.
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S0000685; FIG2.
DQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

```

Query Match	4.7%;	Score 226.5;	DB 1;	Length 1609;
Best Local Similarity	21.1%;	Pred. No. 4.8e-05;		
Matches 206;	Conservative 130;	Mismatches 363;	Indels 279;	Gaps 43

OY 11 STFAIFPLSMI--ATETVLSSASFQDNKNGNFVREOEDACTITYLFKNVTLENIIPG 67
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 362 ATSLDPISSVGSTASFFVGISTANFNQGNINV-VPESTAGSSQYQDMSSSLPLSQT 420

Qy	68	TGTAITKSCFENETGDLTFEGNQN--SLFQYVADATVGAANSSVDKSTFFGSL	125
Db	421	TWVIVNT--NNGQSVTITSPAYVSTAKTYD-GVIEEYTWCPLOTKQALGVS-	475
Qy	126	SFIASPGSITTKGAVSCSTGSLTKNV-----SLSKNF-----	163
Db	476	SISSPAPASFPSSGSSILSSMSSTLAASNINPESTAGSSOYDWMSSSLPQTMWVIN	535
Qy	164	STDNGAIFATK-----LSLQTMALFSENT	191
Db	536	TTNNGQSVTITSPAYVSTAKTYDVITIEYTWCPLOTKQALGISSITISATQTSKP	595
Qy	192	SS--KKG-GALQISDALTTGNO-----GEV----	214
Db	596	SSILTLGISTQLQSDA-TFKGFETINTMLTESTSITEPTFSGTSDSPFLCTSEVINAS	654
Qy	215	---SESDNISDSCSAIFTEASYTISNNK-----VSFLDNVTCASSST	256
Db	655	SLSSITPNSSESGSTATTITNTVTFGTSKRPSTSVNPTSAOHSSVNSLTFTSNS	714
Qy	257	TGDMSGAICA--YKTS-----TDRKVLGNGQMLFESNNTSTAGA-----	297
Db	715	TETI--AVISNIHKTSNMDYSLTTQLTQKSLQJVLSTVTTVNGATEYTTWCPAS	771
Qy	298	---IYKKLEIASGCL-----TLFSRNVGCGAPKGAALAIIDSELSLSDS	343
Db	772	SIAYTTSIKYTLTLTEVCSSHSECTPVTYTSVATSTIP--LSTSSSVLSTSVSE	828
Qy	344	G-----DIYELGNTVSTPCGTNRSSIDLTSAKWTALRSAGRAIYEDTIGSSST	397
Db	829	GAKNPAPASEVTINQVSATSEATSTQVSATSATATASESS-----TTSQYST	877
Qy	398	VTVDLAKNET-----PADSALOYTGNIITFGEKLSLETEAADSKNLSKLOPYT	446
Db	878	ASEIISLIGQNFETTTGSLFPALSTEMINTVTVSRRKLIISTEVCSSHCKVPVITEV	937
Qy	447	LSGGTSLKHK-GVTLQTOAFETQOADSRLBMDVGTLEPADTSTINNVLI-----	494
Db	938	TSKTPENGHSSQJLOLEANEVTLSSHQVIMKT--EVCNSNICPVTYSQMSSTFP	995
Qy	495	-----NISSIDAKKAKIETKATSKNLSLGTITLLDPGTFFYENHSLRNPQSYD--IL	546
Db	996	YLTSSTSSSSSLAKKKSLE--ASSEMTFSVSTOSLPLATFCSEKRSTSVSOWMSNVL	1053
Qy	547	E---LKSGTYTSAATVPDIME-KEHGYO--GTWCPVWGCTAGSTATTNNMKTGY-	599
Db	1054	TNTMSSSSNVTISNEKPSSTPYNFSSGYSLPSSTPSQISLSTAIT-TIMGKITVYT	1112
Qy	600	---IPNPR-----IGSLVENSILMAFIDISLHYMETAN--EGLOQDRAFW	642
Db	1113	TWCPLAEKSTYMASSQSRVDHFVSSKRPSSLSQTSIQYTLSTATTISGLKATVYTTW	1172
Qy	643	CAGLSNFHFKDSTIKTRGCFHLSGCVYIGCNLHTGD--KLSAAPC-----OLF	690
Db	1173	CP-----LTSKSTLDAITQTSSTAKVARTTSASSATISLSISTE	1213
Qy	691	GRDRDYEVAKNOGVYGGTLYUOHNETYISLPCKLPCSLVAPFEIPLFSGNLST--TH	749
Db	1214	SESSSGYLSNG---VCSGTCICDQDVPTQSSSPAS---TLAISP--VSTSSSSSFST	1263
Qy	750	TDNDLTKTKYTTTPVKGS 767	
Db	1264	TASTLTSTHTSVLPBPS 1281	

RESULT 1	
YDBA_ECOLI	STANDARD: PRT: 2003 AA
ID YDBA_ECOLI	
AC P33666; F76087; P76086; F76656; F76857; F76859	
DT 01-FEB-1994 (Rel. 28, Created)	
DT 01-OCT-2000 (Rel. 40, Last sequence update)	
DT 01-OCT-2000 (Rel. 40, Last annotation update)	

CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AB003681; BAA02038.1; -
CC KM Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1 1338 32 KDA BETA PEPTIDE.
CC FT CHAIN 1 1339 1656
CC FT DOMAIN 528 533 POLY-GLY.
CC FT SEQUENCE 1656 AA: 168097 MW: 31326969CD5999F CRC64;

Query Match 4.7% Score 223.5; DB 1; Length 1656;
Best Local Similarity 20.0%; Pred. No. 7.4e-05;
Matches 186; Conservative 113; Mismatches 336; Indels 295; Gaps 44;
QY 22 ATETVLDSSASFQDNKNFSVRESEODACTYLLFKGNVLE-NIPGTATKSCFNN 80
DB 386 ASIVAITQNSFGTDPDGNLAQVTPD--TMTLTGNFTGDANNPQ-----NT 431
QY 81 KQDLTETGNSLFLQVVDAGYVGAAYNSVVDKSTTFGFSLS-----FIASPGS 133
DB 432 AGVITPAANT-----LASASADANAATNNITAEASGVQVLSGTHTAELRLGAGS 486
QY 134 -----SITGK-----GAVSCSTGSLSTFNVSLEFSKPNSTNGG--A 170
DB 487 VEKLADGTIVNGVQTVLVGVLAGAIRLD-GSATITTDIG-----NGGGGA 535
QY 171 ITAKTSLTGTTMSALFSENTSSKKGAIQ--TSDALITTGNGEVSF-----SDNT 220
DB 536 LOSITLANTDKTLTLTGAGNIISANGTINFOANGTIRKLTQNNIVYDCDLAIATDQ 595
QY 221 SSSGGAIFFEASVTISNNKVSFIQNKVQ-----ASSSTGMSGGAIQ----- 266
DB 596 GAVDASSLINAQVLTISGTLIGIIGANNITLGFQFNIGSSKTT--LNGGVAINELVIGNNG 653
QY 267 ---AKTSTDTKVT-LTGNOMLFS--NNTSTAGAIYVKLELASGGLTLEFSNSV 318
DB 654 SVGFANHTYLTFTTNAAGCKIIFNPVNNNTTLAGT-----NLGSAANLAEINFG 707
QY 319 NGSTARK-----GGAIAIEDSGELISADSGDVEL--GNTVSTTPG-----TNBSST 365
DB 708 SKARADTVLVNVEGVNLVATNITTTDANGSFVFNAGGNIVSGVGGQGNKRFVTAL 767
QY 366 DLTGSKMFLRFAAGRAIFYDPTTGSSTVTVTLKVNTPADSLQYTNIFTEGK 425
DB 768 DNGITRK-----FLGNATPFGNTTIA-----ANSTLQJSGN--YTADF 803
QY 426 LSETEA-----DSKMLTSKLLQPYTLG-----GTLSEKHGVTLQ 461
DB 804 IASADGTVIEFVNTGPIVNTLKNQAVPVNALQITVSGGVNVEISAGNVHGAMTD 863
QY 462 TOAFPTQADSRLEM-----DVGTTLEPADSTINNLV-----INISID----- 500
DB 864 TIAFENSSIGAVFLPBGIPFNAGNTIPLTIKSTVGENETAEFSPVSVYSGVDYIAD 923
QY 501 ---GAKKAKIEFKATSKNLTLSGTTLLDPTGTFEENHSLRNPQSDYLELKASGTVS 556
DB 924 GGVIGQNNITVGLGSDNGIIVNATTLVAGIGTINN-----OGTVTL 967
QY 557 TAVTPDPIKGEKHYGQGTWGPVWGCGASTATFNMTKTYIIPPERIGSLVPSLWN 616

DB 968 SGGVNPT-----PQTVYGLGTGIGASKFKOVFT-TDY-----NNLGINIATN--- 1009
QY 617 AFIDISLHYMETANEGLQ-----GDRAPWAGL-----SNFHHKD----- 653
DB 1010 -----TTINDGVTVTTGGTAAAGGIAGTDFDGTITLGSVNGNANVPADGIFSN 1057
QY 654 -----STYTRRG-FRHLSGYVIGGNLHTCSDKILSAFCOLFGFRDYPFAKNQGTIV 706
DB 1058 STSMVTTTANNGTVTYLGNAPV--GNIDSDTPVASVR-----FTGSNNAGL 1104
QY 707 GGLTYQANE--TYSLPKRLAPCSLSTYPTLPVLFSSNLSTHTNDLKTRTYTPPV 764
DB 1105 KGNIVQVDFGTG-----NLGIVNSN--VILGSGSTAINKIDLLTNTLTFAGG 1152
QY 765 KSGMND-----SFALEFGRAPICDESA 789
DB 1153 TSTWGNNTSIETTLTLANGNIGHIYVABGA 1182

RESULT 13
ID EGT2_YEAST STANDARD; PRT; 1041 AA.
AC P42835;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 35, Last annotation update)
DE EGT2 PROTEIN PRECURSOR (EARLY G1 TRANSCRIPT 2).
GN EGT2 OR YNL327W OR NO320.
OC Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1676;
RX MEDLINE: 95373280.
RA Matfah M., Nicaud J.-M., Levesque H., Gallardin C.;
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV
RT identifies the RPD3, PAS8 and KRE1 loci, five new open reading
RT frames.";
RL Yeast 11:567-572(1995).
RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX MEDLINE: 96251274.
RA Kovacech B., Nasmyth K., Schuster T.;
RT "Eg2 gene transcription is induced predominantly by Swi5 in early
RT G1.";
RL Mol. Cell. Biol. 16:3264-3274(1996).
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL
CC SEPARATION AFTER CYTOKINESIS, AS SEPARATION OF MUTANT DAUGHTER
CC CELLS IS DELAYED. COULD EITHER BE AN ENZYME NECESSARY FOR GLUCANS-
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY
CC G1; INACTIVATED BEFORE CELLS PASS START.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: Z46259; CA86371.1; -
CC DR EMBL: Z71603; CA96259.1; -
CC SGN: S0005271. EGT2.
CC Glycoprotein; Repeat; signal; Cell cycle; Cell division; Mitosis.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1041 EGT2 PROTEIN.
FT DOMAIN 200 203 POLY-SER.
FT DOMAIN 381 384 POLY-SER.
FT DOMAIN 388 395 POLY-SER.

```

FT DOMAIN 490 493 POLY-SER.
FT 586 589 POLY-THR.
FT REPEAT 457 492 1-1.
FT REPEAT 577 606 1-2.
FT REPEAT 613 647 1-3.
FT REPEAT 716 745 1-4.
FT REPEAT 773 802 1-5.
FT REPEAT 811 840 1-6.
FT REPEAT 849 886 1-7.
FT REPEAT 887 924 1-8.
FT REPEAT 925 962 1-9.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1041 AA; 108494 MW; 01FECF8BA8744CD CRC64;

```

Query Match 4.7%; Score 222; DB 1; Length 1041;
 Best Local Similarity 23.2%; Pred. No. 4.7e-05;
 Matches 163; Conservative 99; Mismatches 280; Indels 162; Gaps 31;

```

4 SPKPFVFFALPISLMATET-----VLDSASFQGNKNGNFSVRESOE 48
395 SAPSSNSFTTPSSLSATEYISISASISVQASIDNSTTAVTOSTIAVSSAE 454
49 DAGTYLFGKGNVLENIPTGTAITKSCFNNTKGDLLFTGNGSLLFQTVDA----GTV 103
455 KLSSTLSTYNSVTVISVSATQHTTTPSYNSPT-----TLSSSVLEVSISSPYLANTTV 509
104 AGA-----AVNSVVDKSTTFIFGSSLFIA--SPGSSITTGKGVASCCTGSLTKNVS 156
510 SGASASASQSTNPYPYNSNTT---SSATQLATIAPPAINITGISISSITNTSVSSTTS 565
157 LFSKNSFTDNGCAITAKTILTLGTMSALFSENTSSKKGAIQTSDALITIGNOGEVSF 216
566 SISSGFPVNT--AVASGYIILTTTESAQLT-----IGSLIPITITTTTSTTS---- 614
217 SDNTSSDSGAJTEASVATISNNAKVS---FIDNKVTGASSS-----TTGMSGCAICAY 268
615 TDKTGSNK--VASSTELIAQSVNNSSLSVSTINTNATAAANRATFTAHSHGSLQPSY 673
269 KTST-----DTKVTILTGOMLFSNMTSTTAGAIYVKKLELASGILLFNSNNGST 322
674 HSSSLSTIDTKVT-----TATTSISRDGS-----SSLAFTGLNQSVVSTGT 716
323 APKGAIAIEDSGELSLSDSGDIVLGNTVSTPGTKR--SSIDGTSAKMT--ALRSAA 380
717 -DKSDTYVISTESAOVFEYDLSLPISTLKPIVVTGTSRNSFTSVSSTKLEIARATDK 775
381 GRAIYFDPTTGGSTTVTD-----VLKNE-----TPADSALOY- 415
776 GDA---YVSISTQSAOVTEYGMPLSTLEPVTYIMSTDESGYFTLTCCTESGQATEYG 832
416 -----TGNIIT--FTGEKL-----SETEADSKNLISKILOVTLSCGTLISKHGVTLQ 461
833 SLIPISLIDGSAVIYFTGESVVGSTYVGAOYAOHTSLVPAVSTIKGTSLSL----- 885

```

```

OY 462 TQAFTHQADSRLEMDVGTLEPADFTSTNNLVINISSIDGAKKAKIETKATSKNLTLSGT 521
Db 886 -----STEEVSVAAGYITTYGAOYAOHTSLV-PVSTIKRSK-----TSLSTESVAGY 933
OY 522 ITLLDPTGFYENHSLRNPQSYDILELKA--SGTVSTAVTPDPPEKFNHYGYGTWGP 579
Db 934 STYVD--SAQYAEHT--NLVAIDTLKSTFQKATREVCYCTALSSPHSATLIDAGTTIS 989
OY 580 IYWGTCAGTATFNTWTKTYLPNE-----RIGSLV 610
Db 990 LPTSSSTLSLTIITWSSSTIKPSPISFGAGAGQLIRIGSL 1033

RESULT 14
AGAL YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR.
GN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91304412.
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT "The AGAL product is involved in cell surface attachment of the
RL Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RN [2]
RP SEQUENCE FROM N.A.
RA Poll T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S. CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60590; AAA34382.1; -
CC EMBL; Z71659; CAA96325.1; -
CC PIR; S17031; S17031.
CC PIR; A41258; A41258.
CC SGG; S0005327; AGAL.
CC Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
CC Phenomone response.
CC SIGNAL 1 22
CC CHAIN 23 725
CC FT 53 493
CC FT REPEAT 53 149
CC FT REPEAT 395 493
CC FT DOMAIN 182 307
CC FT REPEAT 182 188
CC FT REPEAT 189 195
CC FT REPEAT 196 202
CC FT REPEAT 203 209

```

```

FT REPEAT 210 216 2-5.
FT REPEAT 217 223 2-6.
FT REPEAT 224 230 2-7.
FT REPEAT 231 237 2-8.
FT REPEAT 238 244 2-9.
FT REPEAT 245 251 2-10.
FT REPEAT 252 258 2-11.
FT REPEAT 259 265 2-12.
FT REPEAT 266 272 2-13.
FT REPEAT 273 279 2-14.
FT REPEAT 280 286 2-15.
FT REPEAT 287 293 2-16.
FT REPEAT 294 300 2-17.
FT REPEAT 301 307 2-18.
SQ SEQUENCE 725 AA; 73353 MW; 70420C953B0B01F8 CRC64;

Query Match
Best Local Similarity 22.5%; Score 220.5; DB 1: Length 725;
Matches 135; Conservative 108; Mismatches 269; Indels 89; Gaps 22;

QY 3 SSGPKRV---FSTFAIFPLSMIAETVLDSASFQGNKNGNFVRESQEDAGTYLFGKN 59
   | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
D6 126 SKFTSYICPTCHTAISLSEVGTITVSSSA-IEPSASIIISPVSTLSSTS---SN 180
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 60 VLENIPTGCTA-----ITKCFNNTKGLFTGNGSLFQTVDACTVAGAAV 108
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 181 PTTSLSSSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPST 240
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 109 NSSVVDKSTTFIFGSSLSFASPGSSITTKGAVSCSTGSLTKNVSLFSGKSTNDG 168
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 241 SSSSTSTGSSSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPST 297
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 169 GATTAKTSL-----TGTMSALFSENTSKKGAIQTSALITTGNGSEVFSNDTSSD 223
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 298 PSTLSSPTLASTPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPSTSPST 357
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 224 SGAAIFTEASV--TISNNAKVSFIDNKVTGASSTT-----GDMGCAI-----C 266
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 358 VAPSMSTSVETVSSSSSEI-----TKSISTITPFSMSSTFTTVSGVTMTWTWC 413
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 267 AVKSTDTKVTLLGNOMLFNSNTSTAGCAIYVKKLELASGGLTFERN-SYNGSTAPK 325
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 414 PYSESETS-TLTHSMHEVTDTATCTHESCMPSQTTSLTSSIKKSTKVAVSVSTV 472
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 326 GGAIAIEDGELSADSGIIVFLGNTVISTPCTNRSSIDLQTSAKMPLAKSAGRAIT 365
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 473 ESSVACSTCAETSHYSVQTAASSSVTQOTT--STRKSWVSMSTDEDEFNKHAIRK--- 527
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 386 FYDPITGSGSTVTVDLKVNETPADSALQYTNIFTGKLSFTEADSKNLSKILQPV 445
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 528 -YHVTSSGISTIT---SVSEATSSID-----SESQDSSHLSTSVLSSS 571
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 446 TLGGLSLKHGVLTQTAFTQOADSRLQEMDVGTTLEPADTS-----TNNLVINIS-SI 499
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 572 SLS-ATLSSDSTLL-----FSSVSSLSVEQSPVTTLIQISTSEILOPSTSTAIATISAST 626
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 500 DGAKKRIETKATSKNLTLSGTTLLDPGCTFEHNSLRKPOSDYDLLEKASTVSTAV 559
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 627 SSLSATISIPSTSVESTIES--SLPLPVSSIFLSSSSAPSS-----LOTSVTTTEVST 679
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 560 T 560
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 680 T 680
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

RESULT 15
YH96_YEAST
ID YH96_YEAST STANDARD: PRT: 1140 AA.
AC 004893;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)

```

```

DE HYPOTHETICAL 113.1 KDA PROTEIN IN PR55-FET4 INTERGENIC REGION.
GN YMR317W OR YMR924.09.
OC Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972.
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -I- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 254141; CAA90835.1; -.
DR SGD: S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match
Best Local Similarity 21.2%; Score 218.5; DB 1: Length 1140;
Matches 138; Conservative 107; Mismatches 286; Indels 119; Gaps 20;

QY 18 LSMIATET-----VLDSSASPDGKNGNFVRESQEDAGTYLFGNVTLENIPOTGA 71
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 129 LSSSTLETISISSAIAQTSIPOTSSNGGCS--SEPLGKSSVLETTASSSDT-----TA 181
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 72 ITKSCFNNTKGLFTGNGNSLFPQVDACTVAGAA--VNSSVVDKSTTF----- 119
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 182 VTSSTF-TLLTVSSSPKLISSGSAVTSVGTTSDAKVEFSSSTDVSSLLSTSPASS 240
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 120 -----IGFSS--LSFIASGSSITTKGAVSCSTGSLTKN-----VSLFSKMF 163
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 241 TISELTFSSSTLSTSSVSSSEAPATSSVSSSEASSSTSSVSSSEAPLATSSVSSRA 300
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 164 SPDNGCAIFAKTSLTGTMALFSENTSKKGG-----GAIQTSALITTGNGEVEF 216
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 301 PSTSSVSSSEAPSSSTSSVSSSEISTSSVSSSEAPLATSSVSSSEASSVSSSEI 360
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 217 SDNTSSDGAIFTEASVTISNNAKVSFIDNKVTGASSTGDMGCAICAYKTSTDKV 276
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 361 SSTSSVSSSEAPLATSSVSSSEAPSSSTSSVSSSEAPSSSTSSVSSSEI 420
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 277 TLGNOML-----LFSNNT-----STTAGCAIYVKKLELASGGLTFERNVNGCT 322
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 421 SSTKSSVMSSEVSATSSVSSSEAPSAISSLASRFSKNTSVSTLVATEASSVSSL 480
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 323 APRGCAIAIEDGELSADSGDIV-----ELGNTV-----TSTPGTN 361
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 481 RPSETSLANSIIESLSGYNSTVSTTSAASSTLGKSVSSNSMAMATSKTSSTSDS 540
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 362 RSSIDLGTSAKMTALRSA-----AGRAIFYDPITGSGSTVTVDLKVNETPAD-SALOY 415
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 541 KSVIFGNSSTVTTSPSASISLTASPLPSWMDIITSESSASSISSNLAASSAPSDNNSTIA 600
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 416 TGNLIFTGKLSFTEADSKNLSKILQPVTLGGLSLKHGVLTQTAFTQOADSRLQEM 475
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 601 SASLIVTKTKNSVSSIVSITSETTNSNLATSTSLSKKATARSISTSNATASASAV 660
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 476 DVGIT-----TEPADSTINNVLIN-----ISIDGAKKRIETKATSKNLTLS-- 519
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 661 PGTGFSMSHSTVTPGTSTASAIANSIVSSSLAGSFTSPRESSPTTSTLVTSSEAP 720
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 520 GTITLLDPGCTFEHNSLRNP-----QSDYDLLEKASTVST 557
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
D6 721 STVSSMTTSAPINNSTARSPPSTASTFTSTSSISSVPL-ASGDVITS 769
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

Fri Nov 24 13:50:14 2000

us-09-428-122-2.rsp

Page 14

Search completed: November 20, 2000, 11:33:09
Job time: 722 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 11:20:05 ; Search time 20.34 Seconds
(Without alignments)
4260.138 Million cell updates/sec

Title: US-09-428-122-2
Perfect score: 4774
Sequence: 1 MKSSEPKVFSTFAIFPLSM.....FELRGSSRRYNDLGAKYOF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3865	81.0	746	2	09RB71 Chlamydia p
2	2058	43.1	928	2	09S6P2 Chlamydia p
3	2000	41.9	914	2	086163 Chlamydia p
4	1982	41.5	928	2	09Z398 Chlamydia p
5	1965	41.2	936	2	09Z898 Chlamydia p
6	1940	40.6	930	2	09RB66 Chlamydia p
7	1936	40.6	930	2	09Z393 Chlamydia p
8	1915	40.1	926	2	09Z135 Chlamydia p
9	1855	38.9	928	2	086164 Chlamydia p
10	1656.5	34.7	839	2	P77792 Chlamydia p
11	1641.5	34.4	847	2	P71132 Chlamydia p
12	1604.5	33.6	846	2	P71133 Chlamydia p
13	1573	32.9	841	2	09Z3A1 Chlamydia p
14	1442.5	30.2	922	2	09Z9G5 Chlamydia p
15	1430.5	30.0	922	2	09Z4H9 Chlamydia p
16	1383.5	28.0	1407	2	09Z899 Chlamydia p
17	1377.5	28.9	973	2	09Z896 Chlamydia p
18	1246	26.1	712	2	09RB73 Chlamydia p
19	1148	24.0	458	2	09RB65 Chlamydia p

20	1120.5	23.5	1013	2	084879 Chlamydia p
21	935.5	19.6	649	2	P71134 Chlamydia p
22	923	19.3	354	2	09RB69 Chlamydia p
23	863	18.1	878	2	084882 Chlamydia p
24	821	17.2	494	2	09RB68 Chlamydia p
25	792	16.6	427	2	09RB70 Chlamydia p
26	768	16.1	177	2	09RB72 Chlamydia p
27	711.5	14.9	298	2	09RB67 Chlamydia p
28	685.5	14.4	1609	2	09Z6U5 Chlamydia p
29	685.5	14.4	1609	2	09RB58 Chlamydia p
30	672.5	14.1	978	2	09Z895 Chlamydia p
31	667.5	14.0	978	2	09RB63 Chlamydia p
32	666	14.0	947	2	09Z813 Chlamydia p
33	666	14.0	1723	2	09Z812 Chlamydia p
34	666	14.0	1723	2	09RB59 Chlamydia p
35	665.5	13.9	946	2	09Z880 Chlamydia p
36	665.5	13.9	946	2	09RB60 Chlamydia p
37	646	13.5	934	2	09Z882 Chlamydia p
38	635	13.3	279	2	09Z5Q4 Chlamydia p
39	603.5	12.6	938	2	09Z883 Chlamydia p
40	603	12.6	975	2	084417 Chlamydia t
41	597	12.5	1016	2	084880 Chlamydia t
42	576.5	12.1	1531	2	084818 Chlamydia t
43	559	11.7	1751	2	084419 Chlamydia t
44	523	11.0	1770	2	084418 Chlamydia t
45	513.5	10.8	311	2	09RB64 Chlamydia p

ALIGNMENTS

RESULT 1
09RB71 PRELIMINARY: PRT: 746 AA.
ID 09RB71;
AC SHIRAI M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Itoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kuhara S.;
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-3.2.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138.
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Itoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kuhara S.;
RT "Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and US."
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB033795; BAA85948.1; -
DR EMBL: 746 AA; 80272 MW; 516DE0ED717AF965 CRC64;
SQ SEQUENCE 746 AA; 80272 MW; 516DE0ED717AF965 CRC64;

Query Match 81.0%: Score 3865; DB 2; Length 746;
Best Local Similarity 100.0%: Pred. No. 1.8e-209; Indels 0; Gaps 0;
Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

183 MSALFSESTSSKKGALDTSALITTTGNGEVSFSDNTSSDGAIFTEASVTSSNNAKV 242
|||||
1 MSALFSESTSSKKGALDTSALITTTGNGEVSFSDNTSSDGAIFTEASVTSSNNAKV 60
Db 1 MSALFSESTSSKKGALDTSALITTTGNGEVSFSDNTSSDGAIFTEASVTSSNNAKV 60
OY 243 SFIDNKVTGASSSTTGDMGGAICAYKSTPTKYLTVGNOMLLSSNNSTAGGAIYKK 302
Db 61 SFIDNKVTGASSSTTGDMGGAICAYKSTPTKYLTVGNOMLLSSNNSTAGGAIYKK 120
OY 303 LELASGGLTFSRNVSNGTAPKGAIAIEDSGELISADSGLIVFLGNTVSTPTGTR 362
Db 121 LELASGGLTFSRNVSNGTAPKGAIAIEDSGELISADSGLIVFLGNTVSTPTGTR 180
OY 363 SSIDLGTSAKMTALRSAGRAIYFDPTTGSSTTVTDVLKAVNETPADSALQYIGNIIFT 422

```

|||||
Db 181 SSIDGTSKMTALRSAGRAIYFDPIITGSSSTVTDVLKNETPADSALQYTGNIPT 240
423 GELSTSEADSKNLSKLLQPTLSGGLSLKHGVTLOTOAFTOQADSRLEMDVCTTLE 482
Db 241 GELSTSEADSKNLSKLLQPTLSGGLSLKHGVTLOTOAFTOQADSRLEMDVCTTLE 300
483 PADTSTINNLVINISSIDGAKKAKIETKATSKNLTSGLTLLDPGTGFYENHSLNPOS 542
Db 301 PADTSTINNLVINISSIDGAKKAKIETKATSKNLTSGLTLLDPGTGFYENHSLNPOS 360
543 YDLELKASGTAVSTAATPDPIGKEKHFHYOGCTWGPPIWGTGASTATATFNMKTGYIPN 602
Db 361 YDLELKASGTAVSTAATPDPIGKEKHFHYOGCTWGPPIWGTGASTATATFNMKTGYIPN 420
603 PERIGSLVPSRLNNAFIDISSLHYMETANEGLOGDRAPFACALSNFPHKDSIKTRGR 662
Db 421 PERIGSLVPSRLNNAFIDISSLHYMETANEGLOGDRAPFACALSNFPHKDSIKTRGR 480
663 HLSGGYVIGGNLHTCSDKILSAFCOLFGDRDPYFAKNNGTGYGGGLYYOHNETYISLP 722
Db 481 HLSGGYVIGGNLHTCSDKILSAFCOLFGDRDPYFAKNNGTGYGGGLYYOHNETYISLP 540
723 CKLRPCLSYVPTLEIPVLEFSGNLSTHTNDLTKYTTYPTVKGSGWGNDSFALEFGGRAP 782
Db 541 CKLRPCLSYVPTLEIPVLEFSGNLSTHTNDLTKYTTYPTVKGSGWGNDSFALEFGGRAP 600
783 ICIDSEALFEQYMPFKLOFYVAHOGFKQGTAREFGSSRLVNLALPIGIFEDKESDC 842
Db 601 ICIDSEALFEQYMPFKLOFYVAHOGFKQGTAREFGSSRLVNLALPIGIFEDKESDC 660
843 ODATYMLTIGYTVDLVRSNPDCTTILRISGDSKMTFGTNARQALVLRAGNHCFPSNFE 902
Db 661 ODATYMLTIGYTVDLVRSNPDCTTILRISGDSKMTFGTNARQALVLRAGNHCFPSNFE 720
903 AFSQSFELRGSSRNYNVDLAKYQF 928
Db 721 AFSQSFELRGSSRNYNVDLAKYQF 746

RESULT 2
0956P2 PRELIMINARY: PRT: 928 AA.
AC 0956P2:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR.
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310:
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkegaard S.;
RA "Chlamydia proteins containing the GNAI-repeat belong to a subfamily
RT of autoantigenic pathogenicity factors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ133034; CAB37071.1; -.
KW Signal.
FT SIGNAL.
SO SEQUENCE 1 20 POTENTIAL.
928 AA: 97229 MW: 0590D5206AIDOE1 CRC64:

```

Query Match 43.1%; Score 2058; DB 2: Length 928;
 Best local Similarity 46.7%; Pred. No. 6,9e-108;
 Matches 444; Conservative 142; Mismatches 320; Indels 44; Gaps 20;

```

1 MKSFPKFFV-STFAIF---PLMIATETVLDSSASFDGKN-GNFSVRSQDAGTTL 55
1 MKSQFSLVSLTACFTSCSTVFAATVNIENIGSDSPDSTGTGTYPKNT--TTGIDYT 58

```

```

56 FKNVYTLNIPGTGTAITKSCFNNTKGDLTETGNGNSLLFQTVADGVAAGAVNSVYDK 115
59 LFGDITLQNL-GDSALRKCFSPDSTTESISFAGKGYSLFPLNKS-SAGEGAL-SVITDK 115
116 STTFIGFSSLSFIABGSSITF--GKAVSCSTGSLSLKVNLSLFKSNFSTINDGALTA 173
Db 116 NLSLTGFSSLTFLAAPSIVITPSSGKAVKCG-GDLTDDNNGTILFKODYEENGAIST 174
174 KTLSTLGTMSALFSEPNSTSS---KKGAIQTSALDALTITNGOEVSFSDNTSSDGAALIT 230
Db 175 KNLSTKNGSTGSLSEFGKSSATGKKGALCATGYDIINNTAPLTFSNINIAEAGALINS 234
231 EASVTLISNNAKVSFIDKNKYTGASSSTTGDMGGAICAYKTSTDKVTLTGNOMLFNSNT 290
Db 235 TGNCTITGNTSLVFSENVY---AFAG--NGAL-----SGDADVTISGNQSVTFSGNQ 283
291 STTAGGAIYKKLELAS--GGITLFSRNSVNGTAPKGAALIEDSGELSSADGDIVF 348
Db 284 AVANGAIAIYAKKLLTLAGSGGGISFSNNIVQGTGNGGALISILAGDCSLSAEGDITF 343
349 LGNTVATSTP-GTNRSSIDGTSKMTALRSAGRAIYFDPIITGSSSTVTDVLKNET 407
Db 344 NGNAIYATTPQTTKRNSIDIGSTAKITNLRALSGHSIFFPDITANTADSTDLNLNKA 403
408 PADSAIQYTGNIITFEKLTSEADSKNLSKLLQPTLSGGLSLKHGVTLOTOAFTQ 467
Db 404 DAGNSTDYSGSLVFESEKLTSEDEAKVADNLSTLKQPTLITAGNLVLRKGYLDTKFGTO 463
468 QADSRLEMDVCTTLEPA-DSTINNLVINISSIDGAKKAKIETKATSKNLTSGLTLLD 526
Db 464 TAGSSVIMDAGTTLKASFEVLTGLSIPVDSLGKKVLAASAASKNVALSGTILLD 523
527 PGTGFYENHSLNPOSYDLELKASGTAVSTAATPDPIGKEKHFHYOGCTWGPPIW--- 582
Db 524 NQGNAAENNDLKTQDFSVQSLAGTATTTDVPAPVATPTHGYOGTWC-MTWVDDT 582
583 -GTGASTATFNMKTGYIPNPERIGSLVPSLNAFIDISSLHYMETANEGLOGDRAP 641
Db 583 ASTPKTKITATLMTNIGYLPNPEROGPLVPNSLWGSFSDIOAIOGVIERSAULTLSDROF 642
642 WCAGLSNFFHKDSTKRGRFRLSGGYVIGNLHSCDRIILSAFCOLFGDRDPYFAKN 701
Db 643 WAGVAVNPLDKDKGKKRKYRHKSGGYAIGAAQTCSENLISFACQLFGSKDPLVAKN 702
702 OGTVYGTLYOHNETYISLPCKLRPCLSYV---TEIPVLEFSGNLSTHTNDLTKY 758
Db 703 HTDTYGAAYIGH---ITECGFLICLLDKLPGSMXHPVLVEQLAVSHVSNLTKY 758
759 TTYPTVKGSGWGNDSFALFEGGRAPICLDESALFEQYMPMKLQFYVAHOGFKQGTARE 818
Db 759 TAYPEVKGSGWGNNAFMMLGASHSYPEYLHCFDYPAPYIKLNTLYIRDDSPSEKGTEDR 818
819 EGGSSRLVNLALPIGIRPKESDCODATYNLTLGYTVDLVRSNPDCTTILRISGSKMTE 878
Db 819 SFDSNLFNLSLPIGVKFEKESDCNDSTDLTLSTVPLDIRNDPCTTALVLSGASWET 878
879 GTNLARQALVRAGNHFCFNSNFEAFSOPSFELRGSSRNYNVDLAKYQF 928
Db 879 ANNLARQALVRAGSHYAFSPMEVYLGQVFEVGRSSRLYNVDLGGKQGF 928

```

```

RESULT 3
086163 PRELIMINARY: PRT: 914 AA.
AC 086163:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]

```


RP SEQUENCE FROM N.A.
 RC STRAIN-CD/CWL-029/VR-1310;
 RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
 RT "Identification of two novel genes encoding outer membrane complex
 associated surface layer proteins in *Chlamydia pneumoniae*."
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A001311; CA004671.1; -
 KW Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.
 FT NON_TER 914 914
 SQ SEQUENCE 914 AA: 95603 MW: 88F33BA680F5E3 CRC64:

Query Match 41.9%; Score 2000; DB 2; Length 914;
 Best Local Similarity 46.3%; Pred. No. 1.2e-104;
 Matches 433; Conservative 141; Mismatches 318; Indels 44; Gaps 20;

OY 1 MKSSPKFVF-STFAIF---PLSMIAETVLDSSASFDGKKN-GNSVRSQEDACTYL 55
 DB 1 MKSOFMWLVSLSTLACTSCSTVFAAENIGPSDSFGSTNGTYTPKNT--TTGIDYT 58
 OY 56 FKNVTELENIPGTGTATKSCFNNTKGDILFTFGNSLILFOTVDAGTVAAGAAVNSVDK 115
 DB 59 LTGDITLQNL-GSAAATKGCFSDTFTESLSFAGKGYLSFLNLIK-SAGEAL-SVTYTDK 115
 OY 116 STTFIFGSSLSFIASPOSSITTT--GKAVSCSTGSLSTKNNVLSFSKSTNDGCAITA 173
 DB 116 NLSLTGGSSSLFTLAAPSSVITTPSGKAVCG-GDLTFDNNGLITLFQDQCEENGALIST 174
 OY 174 KTLSLGTGTTMSALFSESTSS---KKGAIOTSDALITNOGEVSEFSDNSSDGAIF 230
 DB 175 KNLISLKNSTGISIEGKSSATGCKKGAICATGIVDTTNNAPLPSNLAEGAAGAINS 234
 OY 231 EASVTSINNAKVSFIDNKKVTAASSSTTGDSGAICAKYKSTDTKYTLTGNNMLFESNNT 290
 DB 235 TGNCITTGNTSLVSENSV---ATAG--NGAL-----SGADVYISGNQSVTFSGNQ 283
 OY 291 STTAGCAIYKKELAS--GGLLFSRNSVNGTAPKGAIAEDSGRLSLADSGIVF 348
 DB 284 AVANGGAIVAKKLLASGGGGGSISSNNIYOGTTAGNGAISLAADKSLSAEAGDITF 343
 OY 349 LGNTVSTTP-GTNRSSIDIGTSAKMTALRSAGRAIFYDPITGTSSTVTDVLKNET 407
 DB 344 NGNAIYATTQTTKRNSIDIGTSKTIINLRAISCHSIFFTPDITANTAASTDTLINKA 403
 OY 408 PADSALOYTNNIIFTGKELSETAADSKNLSKLLQVTLSSGTLKNGVTLQTOAFTQ 467
 DB 404 DAGNSTYSGSIVFSGEKLSEDEKAVADNLTSLKQPVTLTAGNLVKRGVTLDTKGFQ 463
 OY 468 QADSRLMDVGTLEPA-DSTINNLVINISSIDAKKAIETKATSKNLTSLTILLD 526
 DB 464 TAASSVIMDAGTTLKASTETVTLGLSIPVDSIGEGKVVIAASAAKNALSGPILLD 523
 OY 527 PTGTFEYENHSLRNPQSYDIELKASGVTSTAVTPDPIMEKFFHYQGTGPIVW--- 582
 DB 524 NQGNAYENHDLGKTQDSFQOLSAIGTATTTPDAVPATVAPTHYGGQGWG-METWDDT 582
 OY 583 -GTGASTTAFNNKTGYIDNPERIGSLVNSLMAFIIDISLHYLMEFANEGLOGRAF 641
 DB 583 ASPTKLTATLAWNTGYLNNPEROGPLVNSLWGSFSDIOAGVIERALSALTCSDRGF 642
 OY 642 WCGLSLSEFFKDKSTRGRGRHLISGVVIGVGNLHATCSDKLISAFCOLPDRDRYPAKN 701
 DB 643 WAAGVAFELDKDKGKRRKRHSKGAIGAAOTCSENLISFAFCOLPSSDKDFLAKN 702
 OY 702 QGTAVGGLTYOHNETYISLPCKLRPSLSYVP---TEIPVLSGNSLSYHTDNDLTKY 758
 DB 703 HTDTYAAAFIYOH---ITFCSGFICGLDKLPKSGSHKRLVLEGOAYSHVNSDLETKY 758
 OY 759 TTYPTVYKSGNSDFALEFGRAPICLDESALFEQYMPFNKLOFVYAHOGEKFEQTEAR 818
 DB 759 TAYPEVGSNGNNAFMNLGASSHSYPEYLHCFDTYAPYIKLNLTYLRQDSFSEKTEGR 818

OY 819 EFGSSRLVNLALPIGIRFEDKESDCCOATYNLTGYTVLVRNSPDCCTTLRISGDSWKTF 878
 DB 819 SFDSLSLNFSLPIGVKFKFSKDCNDFSYDLTSLYVPLDIRNDPKCTFLVLSGASWEY 878
 OY 879 GTNLARQALVLRAGNFCFNSNFARSQSFELRGS 914
 DB 879 ANNLRQALQVLRAGSHYAFSPMEFVLGQFVEVRGS 914

RESULT 4
 ID 092398 PRELIMINARY: PRT: 928 AA.
 AC 092398;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE OUTER MEMBRANE PROTEIN PRECURSOR.
 GN OMP10 OR PMP-9.
 OS *Chlamydia pneumoniae* (*Chlamydia pneumoniae*).
 OC Bacteria: Chlamydiales; Chlamydiales; Chlamydiales.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE: 99206606.
 RA Kalman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*."
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VR1310;
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christiansen G., Birkelund S.;
 RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily
 of auto-transferring pathogenicity factors."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0138;
 RA Shirai M., Hiraoka H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
 RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
 RA Matsushima H., Tanaka C., Furukawa S., Mura K., Nakazawa A.,
 RA Ishii K., Shiba T., Hattori M., Kohara S.;
 RT "Comparison of outer membrane protein genes omp and pmp in the whole
 genome sequences of *Chlamydia pneumoniae* isolates from Japan and
 RT US."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE001628; AAD18591.1; -
 DR EMBL: AJ133034; CAB37069.1; -
 DR EMBL: AB033803; BAA85956.1; -
 KW Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.
 SQ SEQUENCE 928 AA: 98332 MW: 58910A8F04F1219 CRC64:

Query Match 41.5%; Score 1982; DB 2; Length 928;
 Best Local Similarity 45.2%; Pred. No. 1.3e-103;
 Matches 429; Conservative 150; Mismatches 327; Indels 44; Gaps 19;

OY 1 MKSSPKFVFSTFAIFPLSM-----IATEVLDSSASEFGKNGNFSVRESQEDA-GTT 53
 DB 1 MKSSLMLFLSSSLALPLINFAFAAYEINLGPINSSG--PGTYTTPAOTTNMDGTI 58
 OY 54 YLFKGVNTELENIPGTGTATKSCFNNTKGDILFTFGNSLILFOTVDAGTVAAGAAVNSVY 113
 DB 59 YNLTDGVSTIN-AGSPTALALASCFFKETTGNLSFQGHGYFLLLQNIIDAG--ANCTFTNTAA 115
 OY 114 DKSTTFIGSSLSFIASPOSSITTGKAVSCSTGSLSTKNNVLSFSKSTNDGCAITA 173
 DB 116 NKLLSFGFSYSLIOT--TNATGTGAIRK-STGACISQISNISCYFGNPSNDNGGALOG 172

09R866
ID 09R866 PRELIMINARY: PRT: 930 AA.
AC 09R866:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DE POLYORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP_8.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138.
RA Shitai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kuhara S.:
RT "Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and US";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033802; BAA85955.1; -
SQ SEQUENCE 930 AA: 97639 MW: 46a8896761391C09 CRC64;

Query Match 40.6%; Score 1940; DB 2; Length 930;
Best Local Similarity 43.7%; Pred. No. 2.9e-101;
Matches 416; Conservative 165; Mismatches 324; Indels 48; Gaps 17;

1 MKSFPKFEVSTFAIFPLSM-IAT---ETVLDSSASFQKNGNFVSRESQEDAGTYLIF 56
1 MKPIPKLLISSFLVPIILSTIATYGADASLPTDSFDGAGSGSTFPKSLADANGTYVL 60
57 KGNVLENIPIGDTATKSCFNNTKGDLTFTGNGNSLFTQVADAGVAGAAVSSVVDKS 116
61 SGNVYI-NDAGKGTALTGCCFTETGDLFTGKGSFSTVADAGSAA-STRADKA 118
117 TTFIGFSSLFASPSSITFGKAVSCSTGSLTKRNVLSFKNSPQ---NGGALTA 123
119 LFTGFSNLSFLIAPGTIVASGKSTLS-SAGALNLTDNGILTSQNVSEANNNGALTA 177
174 KTLISLTGTMASLFSENTSSKKGAIQISDALITGNOGEVSESDTSSISGAIFTEAS 233
178 KTLISGNTSITFTSNAKKGAIYSSAASISGVTGLVNNKKGTEGAGLEAS 237
234 VTISSNAKVSFIDNKYTGASSSTTGDMSCGALCAIKTSTKTLTGNGQMLLSNNTST 293
238 SSITQNSSLFEFGSNTATDAG-----KGAIIICEKGETPLTLLISGNKSLTFEANSV 291
294 AGGAIYVKKLELASGGLTFSRNSVNGTTPKKGAIIEDESGELSLADSGDIVFGNTV 353
292 QGGAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIALDAGSLSLSAAGDITFLCNTL 351
354 TSTT-PTNNSSIDLGTSAKMTALRSAGRAIFYDPI---TTGSSITVYDVLKIVETPA 409
352 TSTTAPTSTRNAIYLGSSAKITMLRAAGOSIIFYDPIASNTGAS---DVLINOPDS 407
410 DSALOTGNTIIFGKELSTFEADSKNLTKLQPTLSGTSILKHGVTLOTQAFQQA 469
408 NSPLDGSGLTIFGKELSADEAAADNFSTILKQPLALAGLALGKVELDNGVFQTE 467
470 DSRLEMDVGTLEPADTSTN--NLVINISSIGAKAKETKATSKNLSTGTLTLP 527
468 GGLLMQPTKRL-ADTEALSTKLIVDLSALGKNSVSLETGANKTITLISPLVPODS 526
528 TGTFFNHSILRNPOSYDILELAS-----GYVTSTAVPDPIMGEKFEHYGOTWG 578
527 SCNFYESSHTNQAFQPLVVFATAASDIYIDALLITSPOTEP-----HYGYGHW 580
579 PIWGGASTATFENMTKGTIYNPERIGSLVNSLMNADISSLHYEMETANEGLQGD 638
581 ATYADSTSTASGIMVTWITTYINPERRASVVDLSMASTDIRILOQITSOANSIYOO 640
639 RAFWAGLANSFHKDSTKTRRGEFRLHSGGVYIGNLTGCSDKILSAAFCOLFGRDRFY 698

DB 641 RGLMAGTIANFHHKDSGTNOAFRRHKSXYIVGGSNADESENIFFSAFQGLKDKDLFI 700
699 AKNOGVYCGTLYXONENETIISLPCKLRPC--SLSVPTPEIVLFGNLSYTHNDLKT 756
701 VENTSINYLASLYLQHRALGLSP---MPSFGSTIDMLKDIFILINAQISYTKNDMT 757
757 KYTTYPTVKSGNDSFALEFGRAFICL-DESALEQOYMPFMKLOFYVAHOGEKQGT 815
758 RQSYEPAGOSWTNNGALGSLALYLPKPAFFQGYFPLKQAVYSROONFKESGA 817
816 EAREFSSRLVNLALPIGIRPDKESQDQATYNTLIGYTVDLVRSNPDCTTLRLISGDSW 875
818 EARAEPDGLVNCISIPVIGIRLEKISDEKNNFELISATYIGDYRKRPRTSLMWSGASW 877
876 KTFGTMLARQALVLRAGNHCFNSNFARSQSFELRGSSRYNVDIGAKYOF 928
878 TSLCKNLARQAFSLASGSLTILSPHVELSGEAAVELRGSAAHIYNDGGLRYSF 930

RESULT 7
092393
ID 092393 PRELIMINARY: PRT: 930 AA.
AC 092393:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE OUTER MEMBRANE PROTEIN 11 PRECURSOR.
GN OMP11 OR PMP 8.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029.
RC MEDLINE: 99206606.
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.:
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RT Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310.
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.:
RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily
of autoexporting pathogenicity factors";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE001627; AAD18590.1; -
DR EMBL: AJ133034; CAB37068.1; -
KW Signal.
FT SIGNAL 1 27
FT CHAIN 28 930
FT SEQUENCE 930 AA: 97669 MW: 46A9B5E3BB913C4C CRC64;

Query Match 40.6%; Score 1936; DB 2; Length 930;
Best Local Similarity 43.5%; Pred. No. 4.9e-101;
Matches 415; Conservative 165; Mismatches 323; Indels 48; Gaps 17;

1 MKSFPKFEVSTFAIFPLSM-IAT---ETVLDSSASFQKNGNFVSRESQEDAGTYLIF 56
1 MKPIPKLLISSFLVPIILSTIATYGADASLPTDSFDGAGSGSTFPKSLADANGTYVL 60
57 KGNVLENIPIGDTATKSCFNNTKGDLTFTGNGNSLFTQVADAGVAGAAVSSVVDKS 116
61 SGNVYI-NDAGKGTALTGCCFTETGDLFTGKGSFSTVADAGSAA-STRADKA 118
117 TTFIGFSSLFASPSSITFGKAVSCSTGSLTKRNVLSFKNSPQ---NGGALTA 123
119 LFTGFSNLSFLIAPGTIVASGKSTLS-SAGALNLTDNGILTSQNVSEANNNGALTA 177
174 KTLISLTGTMASLFSENTSSKKGAIQISDALITGNOGEVSESDTSSISGAIFTEAS 233
178 KTLISGNTSITFTSNAKKGAIYSSAASISGVTGLVNNKKGTEGAGLEAS 237
234 VTISSNAKVSFIDNKYTGASSSTTGDMSCGALCAIKTSTKTLTGNGQMLLSNNTST 293
238 SSITQNSSLFEFGSNTATDAG-----KGAIIICEKGETPLTLLISGNKSLTFEANSV 291
294 AGGAIYVKKLELASGGLTFSRNSVNGTTPKKGAIIEDESGELSLADSGDIVFGNTV 353
292 QGGAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIALDAGSLSLSAAGDITFLCNTL 351
354 TSTT-PTNNSSIDLGTSAKMTALRSAGRAIFYDPI---TTGSSITVYDVLKIVETPA 409
352 TSTTAPTSTRNAIYLGSSAKITMLRAAGOSIIFYDPIASNTGAS---DVLINOPDS 407
410 DSALOTGNTIIFGKELSTFEADSKNLTKLQPTLSGTSILKHGVTLOTQAFQQA 469
408 NSPLDGSGLTIFGKELSADEAAADNFSTILKQPLALAGLALGKVELDNGVFQTE 467
470 DSRLEMDVGTLEPADTSTN--NLVINISSIGAKAKETKATSKNLSTGTLTLP 527
468 GGLLMQPTKRL-ADTEALSTKLIVDLSALGKNSVSLETGANKTITLISPLVPODS 526
528 TGTFFNHSILRNPOSYDILELAS-----GYVTSTAVPDPIMGEKFEHYGOTWG 578
527 SCNFYESSHTNQAFQPLVVFATAASDIYIDALLITSPOTEP-----HYGYGHW 580
579 PIWGGASTATFENMTKGTIYNPERIGSLVNSLMNADISSLHYEMETANEGLQGD 638
581 ATYADSTSTASGIMVTWITTYINPERRASVVDLSMASTDIRILOQITSOANSIYOO 640
639 RAFWAGLANSFHKDSTKTRRGEFRLHSGGVYIGNLTGCSDKILSAAFCOLFGRDRFY 698

Query Match	40.1%;	Score 1915;	DB 2;	Length 926;
Best Local Similarity	42.9%;	Pred. No. 7.4e-100;		
Matches 410;	Conservative 156;	Mismatches 333;	Indels 56;	Gaps 18

[illegible]

DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMP4 OR PMP11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC/CW-029/VR-1310.
RA Knudsen K., Madsen A.S., Myind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VR1310.
RA Hjerno K., Boesen T., Dagaard L., Knudsen K., Madsen A.,
RT Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GCAI-repeat belong to a subfamily
of autotransporting pathogenicity factors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CW1029;
RX MEDLINE: 99206606.
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Greenwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takuchi A., Nishida J., Shibata K., Fujiwara R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shida I., Hattori M., Kunara S.;
RT "Comparison of outer membrane protein genes omp and pmp in the whole
RT genome sequences of Chlamydia pneumoniae isolates from Japan and
US."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ001311; CA004672.1; -;
DR EMBL: AJ133034; CAB3702.1; -;
DR EMBL: AE001628; AAD18593.1; -;
DR EMBL: AB033806; BAA85959.1; -;
KW signal.
FT SIGNAL.
FT CHAIN 1 17 POTENTIAL.
FT SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

Query Match 38.9%; Score 1855; DB 2; Length 928;
Best Local Similarity 42.7%; Pred. No. 1.7e-96;
Matches 405; Conservative 171; Mismatches 331; Indels 42; Gaps 20;

QY 1 MMSFFKVFSTFAIRP--LSMATEVYDDSSAFDGN-KNGFVSREQEDAGTYLTK 57
DB 1 MKTSLPWWVSVSLAFSCHLQSLANELLSPPDSFNGNIDSGFTPKTS---ATVYSLT 56
OY 58 GAVTLENIGTGTAIRKSGFNNTKGDLPFTNGNSNLLPQTVAGATAGAAVNSVVDKST 117
DB 57 GGVFFYE-PKGTPLSDSCFKOTTDNLVFLGNHSLTFPGIDGTHAGAA-STANKNL 114
OY 118 TFIGSSLSFLIAPGSSITFGKAVSCSTGSLSTKNVLSFKNSTNDNGAIFAKTIS 177
DB 115 TFSGELLISFSDSPSTVTYTGOTIS-SAGGVLENIRKLVAAGNSTPDGAIKASFL 173
OY 178 LIGTMSALFSEMTSSKKGATOTSDALITITGNOGVSSDNTSSSGAIFTEASVTIS 237
DB 174 LGTSDALFNSNNSSTKGAIAITTAGARIANNVGYRFLSNASTSGAIDDEGTSILS 233
OY 238 NNAKVFINKVTYTGASSTGTGDMGCAICAYKTSTYTLTGNOQLLSNNTSTAGGA 297
DB 234 NKKFLVF-----EGNAKTT-----GGAICNTKASSGPELLISNKKLIRASNAVAETSGA 284
OY 298 IYKKLELASGGTLTFSRNSVNGGTAPKGAIAIEDSGELSLSDSGDIVFLGNTVTSF- 356

DB 285 IIAKKLASGGFTEFLRNWSSAT-PKGAISIDASELSAETGNTITVRNLTFTTG 343
OY 357 -TPGNRSSIDGTSAKMTALRSAGRAIYFDPITTSSTVTVLVYNENPADAQY 415
DB 344 STDTPKRNAINNGSKFTELRAKNHITFFDPTT--SEGSSVLKINNRSALAPY 401
OY 416 TGNITFGKLESEPADSKNLTSLKLOPVTLSGGTSLKHGVTLOTQAFTQADSRLEM 475
DB 402 OCTIFSGEFTLADKLKADNLSSTFOPVSLGGKLLKQKVTLESTFSQEAASLIGM 461
OY 476 DVGTFLE-PADSTNNINVINISIDGAKKAKIETKATSKNLTSGITLTDPTFEFN 534
DB 462 DSGITLSTTAGITTTTNGINVDLSGLKOPVSLAKGASNKYISGKLNLDIEGNITES 521
OY 535 HSLRNPQSYDILELKAAGTVIS---TAVTPDPIGKEFFHYGQGTWGPVWGASTT- 589
DB 522 HMFSDQLFSLKIVDADVDDNDVSSILPVPADPSEYGFQGNV-VNMTTDTATNT 580
OY 590 --ATFNWTKTGIPNPERIGSLVPSLWNAFLDISLSLHMETANEGIQGDRAPWAGLS 647
DB 581 KEATATWTKTGFPSPERSALVCNTLWGVFDIRSLQOLVEIGATGHEHKGFWVSSMT 640
OY 648 NFHKDSTRKRGFRHLSGGVYIGNLHTGSPDKILSAFCOLFGDRDYFAVAKNOGTYYG 707
DB 641 NKLHTGDENKRGFRHNSGGVYIGSANTPKDDLTFEAFCHLFADKCFIAHNNSTYTG 700
OY 708 GTLYVQHNET----YISL-PCKLRCSLYVPEIPVLFSGNLSYTHTDNDLTKTYTY 761
DB 701 GTLPFKSHITLOPQNYLRIGRAKFSESAIEKPREIPALDVOVSFSHSDNMETHYSL 760
OY 762 PVKSWGNSDFALEFGGRAPICL-DESALEQVWPEPKLOFVYAHOGFEGQEAEEF 820
DB 761 PSEBSWSNECIAAGIGIDLPRVLSNPHLKTFTIPQKVEVWYVSONSFESSSDGCGF 820
OY 821 GSSRVNLTALPIGIFEDDESQDQA-TYNLTIGYVADVRSNPDCITLURISGDSKFTFG 879
DB 821 STGRLLNLSIPYGANF-VQGDIGDSYTYDLGFSVDYVRNNPOSTATLVSPDSKIRG 879
OY 880 TNLARQALVLRAGNFCFNSNFARFQSFELRGSSRNINVDLAKKYP 928
DB 880 GNLSRQAFLLRGSNNYVYNSNCELFGHAMELRGSSRNINVDVGTKRF 928

RESULT 10
P77792 PRELIMINARY; PRT; 839 AA.
AC P77792;
DI 01-FEB-1997 (TREMblrel. 02, Created)
DI 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DI 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE POMP908 PRECURSOR.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U65943; AAC15924.1; -;
DR EMBL: U65942; AAC15922.1; -;
KW signal.
FT SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT SEQUENCE 839 AA; 89824 MW; 4581C7CBAF7FF4C4 CRC64;

Query Match 34.7%; Score 1656.5; DB 2; Length 839;

QY	635	LOGRAEACALSNFPHKDSKTRGRGRHSGGVJGNLHTOSDKTISARCOLFGRDR	694
Db <td>548</td> <td>ADYHRGEVSSLGNFHLKSSGSDTKRKRHRNSAGATAGVAYQTSSEDEVTSARCOLFGKDR</td> <td>607</td>	548	ADYHRGEVSSLGNFHLKSSGSDTKRKRHRNSAGATAGVAYQTSSEDEVTSARCOLFGKDR	607
QY <td>695</td> <td>DYFAKNQGTIVGGTGLVYQHNETYISLPCKLRPCSLSYPT-----EIPVFSG</td> <td>743</td>	695	DYFAKNQGTIVGGTGLVYQHNETYISLPCKLRPCSLSYPT-----EIPVFSG	743
Db <td>608</td> <td>DYLVSKNSSTYAGSIYQH-----ISYNTWTLLQNTLGAAPVLNA</td> <td>652</td>	608	DYLVSKNSSTYAGSIYQH-----ISYNTWTLLQNTLGAAPVLNA	652
QY <td>744</td> <td>NLSYTHDNDLKTXY-----ITYTVVGSNGNSFALFEGGRAPICLDSALFQYM</td> <td>795</td>	744	NLSYTHDNDLKTXY-----ITYTVVGSNGNSFALFEGGRAPICLDSALFQYM	795
Db <td>653</td> <td>QLACHASNNKTKMTMTDYABPKTYTSEYIGDNGNDCEFGVEFAKAPI-ETASLFDMS</td> <td>711</td>	653	QLACHASNNKTKMTMTDYABPKTYTSEYIGDNGNDCEFGVEFAKAPI-ETASLFDMS	711
QY <td>796</td> <td>PEMKLOEYVAHQEGFEKGT-EAREFESSRLVNLALPIGIRGEDSQDQATYNTLIGYT</td> <td>854</td>	796	PEMKLOEYVAHQEGFEKGT-EAREFESSRLVNLALPIGIRGEDSQDQATYNTLIGYT	854
Db <td>712</td> <td>PFVKQLQVHANQDQFKEKNSQGGYFEFNNLJNLMSPIGVKLEKESHKPTASVNLTLAA</td> <td>771</td>	712	PFVKQLQVHANQDQFKEKNSQGGYFEFNNLJNLMSPIGVKLEKESHKPTASVNLTLAA	771
QY <td>835</td> <td>VDLYRANPDCTTIRIGSDS--WKTFTNLARQALVLRAGNMFCSNFSAFSQSFELR</td> <td>912</td>	835	VDLYRANPDCTTIRIGSDS--WKTFTNLARQALVLRAGNMFCSNFSAFSQSFELR	912
Db <td>772</td> <td>PDIYRNSDCTASLVSPTSASVWTKANNLRARAFILQAGNYLALRYNELFSQFEFLR</td> <td>831</td>	772	PDIYRNSDCTASLVSPTSASVWTKANNLRARAFILQAGNYLALRYNELFSQFEFLR	831
QY <td>913</td> <td>GSSRYNYNDLAKYQF</td> <td>928</td>	913	GSSRYNYNDLAKYQF	928
Db <td>832</td> <td>GSCRNYNDLSKIQF</td> <td>847</td>	832	GSCRNYNDLSKIQF	847
RESULT	12		
P71133			
ID	P71133	PRELIMINARY;	PRT; 846 AA.
AC	P71133:		
DT	01-FEB-1997 (TREMBLrel_02, Created)		
DT	01-FEB-1997 (TREMBLrel_02, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel_13, Last annotation update)		
DE	POMP91B PRECURSOR.		
OS	Chlamydia psittaci.		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.		
FN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;		
RX	Longbottom D, Russell M., Jones G.E., Lainson F.A., Herring A.J.;		
RX	Longbottom D, Russell M., Jones G.E., Lainson F.A., Herring A.J.;		
RT	"Identification of a multigene family coding for the 90 kDa proteins		
RT	of the ovine abortion subtypes of Chlamydia psittaci.";		
RL	FEMS-Microbiol. Lett. 142:277-281(1996).		
FN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;		
RX	MEDLINE; 98187897.		
RA	Longbottom D, Russell M., Dunbar S.M., Jones G.E., Herring A.J.;		
RT	"Molecular cloning and characterization of the genes coding for the		
RT	highly immunogenic cluster of 90-kilodalton envelope proteins from the		
RL	Chlamydia psittaci subtype that causes abortion in sheep.";		
RL	Embl; Immun. 66:1317-1324(1998).		
DR	Embl; U65943; AAC15923.1; -.		
KW	Signal.		
FT	SIGNAL		
ET	CHAIN	1 16	POTENTIAL.
ET	CHAIN	17 846	POMP91B.
SO	SEQUENCE	846 AA; 90834 MW; 4CDD31DC03C2964E CRC64;	

	Query Match	Similarity	33.6%	Score	1604.5	DB 2	Length	846	
	Best Local	Similarity	38.7%	Pred. No.	1.8e-82				
	Matches	368	Conservative	140	Mismatches	269	Indels	173	Gaps
Oy	23	TETVDDSSASFQGN-KNGNFSYRESOEDAGTYYLFKNGNTLENIPOTGAITKSCFNMTK	81						
		: : : : : : : :							
Dd	26	TNETLTSGDSYGNANTSDFEVKET--TSGAIITYTEGNYCI-SYAKGDSPLKSCSTSETT	82						
Oy	82	GDITFTGNGNSLTFOTVDAGTYAGAAVNSSVVDKSTTFIFGS--SLSFASDPSSTTTTK	139						
		: : : : : : : :							
Dd	83	EHLSTIGNRYTLCPFNITTQSHPGALISVSGINKTLIDISGFLEFCAIYCPCPG---TTY	139						
Oy	140	CAVSCSTGSLTKNVSLFSKNFSTDNGAIIATAKTLSLTGTMGALRENTSKKGAI	199						

```

Db 140 GAIQ-TKOTTTLLKNNSSLVEFKHKNSTAEGLQCK----- 173
OY 200 QTSDALITTTGNOGEVSESDNTSSDGAIAITEASVITISNNAKVSFIIDNKVTGASSSTTGD 259
Db 174 -----SSSSTAE 180
OY 260 MSGGAIICAKRKSTDTKTYLITGNOMLLEFSNNTSTAGAIIVYKLELASEGLITFERSVN 319
Db 181 LK-----LENNKKLIVSESSSKKEKGAIAIDKLITTVSGPTLFENNSVS 224
OY 320 GGTAPKGAIAIEDS-GELSIADSGSIVFNGVST---TPGTNNSSIDLTSAKMTA 375
Db 225 HNSSRKGAICIKRSDGECSLTANLGDITTDGKIIITNGCSPVYRNSIDLSSGKFTK 284
OY 376 LRSAGRAIYFDEPI-TTGSSTTVTDVLYKVENPDAADALQYTGNIITFGKELSETEADS 434
Db 285 LNAKGFIFFYDPIANNGSTEI---ELNKTESDTH-YTGKIVYSGEKLSDKEKTVP 338
OY 435 KNLTSKLLQPTLSGCGTSLSKHGVTLOTQAFITQOADSRLKMDVGTILE---PADSTIN 490
Db 339 ANLKXYFQKPIKIAGSLVLDGVTLEAKKKTQTKRGSTVAMDGLTLOTSSSGEITTL 398
OY 491 NLVIVISSI---DQAKKAKIETKATSKNLTLISGTTILDPDTGFYENHSIRNPOSVDILE 547
Db 399 NLIDINIASLGGGGGTAPAKKLTTNTASQAISIA-AVNLVNTDSMTYEDPIILSASKSFSAT- 456
OY 548 LKASGTVTSTAVTDPDPIMGEXF---HYGYGTWGPVWGTSAT--TATFNWTKTYI 600
Db 457 ---TATTSSTVTPPTENLKNYTPPTHYGYQNN-VTYMKQSSAQDEKTTLWEDQTVS 512
OY 601 PNPERIGSLVNSLMAWAFIDISSLHYMETANBGLQGDRAFCWAGLSNPFHKDSTYTRRC 660
Db 513 PNPERVGSGLVNTLTMGAFSDRAIONLMDISVNGADYSRGFWVSSLANFLNKGSGSTTRK 572
OY 661 FRHLSGTVIGGNLHTCSDKLISAFQQLGGRDPDYVVAANOGTVYGGTLYOHNETYIS 720
Db 573 FRHHSGAGTALVYVQTPSDSDCSAFQQLGKDDYFVSKNSSTIYAGSLTYOH----- 626
OY 721 LPCKLRPCSLSEVPL-----ELPYLFGSGLSYTTHDNDLTKYV-----TY 761
Db 627 -----ISYNTWTNTLLONTLGAERAPLVLAQOLTYCHASNNKMTNMTNTYTPKANVT 677
OY 762 PTVKSGMGNDSFALEFGGRAPICLDESALFEQYWPFMKLOFYVAHOEGFREOGT-EARE 820
Db 678 SEIKGDMNDCEFGVEFGAKAPI-ETASILLFDWYSPFYKQLVHAHQDDFRENNSDGGRXF 736
OY 821 GSSRLVNLALPIGRFDPKESCODATNVLILGTIVDLVRSNPOCTTLTRISGGS--WKTF 878
Db 737 ESNNTJNLSMPIGVKLEKFSHKDASYNLTLAAPDIVRNSNPCOTSLSLSPYSAAVWTK 796
OY 879 GTNLAROLVLRAGNHFCFNSNFEAFQSPSEILRGSSRNYNVDLGAQYF 928
Db 797 ANNLRARAFILQAGNLYALTRNTLTLFSQFQFELRGSCRTYNNIDGSKIOF 846

```

RESULT	13		
0923A1			
ID	0923A1	PRELIMINARY;	PRF; 841 AA.
AC	0923A1;		
DT	01-MAY-1999 (TREMBLE)	10.	Created)
DT	01-MAY-1999 (TREMBLE)	1.	Last sequence update)
DT	01-MAY-2000 (TREMBLE)	13.	Last annotation update)
DE	OUTER MEMBRANE PROTEIN 7.		
CN	OMP7 OR PMP-2.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
NC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OC	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CWL029;		
RX	MEDLINE: 99206606		
RA	Kelman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;		

OM of: US-09-428-122-2 to: GenEmbl:* out_format : pfs

Date: Nov 20, 2000 4:36 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=itame+g2n.model -DEV=xlp  
-o=/cgn2.1/USPTO_spool/US09428122/runat.14112000.163316.27911/app-query.fasta.1.999  
-B=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09428122.@CGN1.1.5773 -NCPU=6  
-ICPU=3 -LONGLOG -NO_XLPPX -WAIT -THREADS=1
```

Search information block:

Query: US-09-428-122-2

Query length: 928

Database: GenEmbl:*

Database sequences: 1033670

Search time (sec): 2334.060000

Score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_pat:AE002235	- 4760.00	4008.25	1.7e-216	2787	AB1835 Sequence 9 from Patent
gb_pat:AE001835	- 4746.00	4002.48	3.8e-215	12127	AE002235 Chlamydia pneumoniae
gb_pat:AE001587	+ 4746.00	4005.73	5.3e-215	16448	AE001587 Chlamydia pneumoniae
gb_pat:CPN133035	+ 4746.00	4001.60	9.0e-215	26920	AJ133035 Chlamydia pneumoniae
gb_pat:AP002545	+ 4746.00	3981.41	1.2e-213	300050	AP002545 Chlamydia pneumoniae
gb_pat:AB033795	+ 3865.00	3276.51	2.2e-174	2241	AB033795 Chlamydia pneumoniae
gb_pat:AB1829	+ 3058.00	1744.59	4.6e-89	2815	AB1829 Sequence 3 from Patent
gb_pat:AE002192	+ 2058.00	1732.09	2.3e-88	12656	AE002192 Chlamydia pneumoniae
gb_pat:CPN133034	- 2058.00	1729.49	3.3e-88	17280	AJ133034 Chlamydia pneumoniae
gb_pat:AP002546	- 2058.00	1705.58	7.0e-87	300600	AP002546 Chlamydia pneumoniae
gb_pat:AE001628	- 2044.00	1721.61	8.9e-88	10757	AE001628 Chlamydia pneumoniae
gb_pat:AB1837	+ 2001.00	1696.60	2.2e-86	2757	AB1837 Sequence 11 from Patent
gb_pat:AB1851	+ 2000.00	1695.05	2.7e-86	3000	AB1851 Sequence 25 from Patent
gb_pat:CPN133034	- 2000.00	1689.20	5.7e-86	6030	AJ001311 Chlamydia pneumoniae
gb_pat:AB1835	+ 1988.00	1683.81	1.1e-85	2787	AB1835 Sequence 13 from Patent
gb_pat:AB033803	+ 1982.00	1680.43	1.8e-85	2787	AB033803 Chlamydia pneumoniae
gb_pat:AE001628	+ 1982.00	1669.12	7.5e-85	10757	AE001628 Chlamydia pneumoniae
gb_pat:CPN133034	- 1982.00	1667.74	9.0e-85	12676	AE002192 Chlamydia pneumoniae
gb_pat:AP002545	+ 1982.00	1665.15	1.2e-84	17280	AJ133034 Chlamydia pneumoniae
gb_pat:AE001627	+ 1965.00	1651.23	2.7e-83	300600	AE001627 Chlamydia pneumoniae
gb_pat:AB033849	+ 1964.00	1655.69	4.2e-84	8658	AB033849 Chlamydia pneumoniae
gb_pat:AE002193	- 1964.00	1654.47	4.9e-84	10026	AE002193 Chlamydia pneumoniae
gb_pat:AB033802	+ 1940.00	1644.85	1.7e-83	2793	AB033802 Chlamydia pneumoniae
gb_pat:CPN133034	+ 1927.00	1633.84	6.9e-83	2793	AJ1841 Sequence 15 from Patent
gb_pat:AB1841	+ 1915.00	1618.93	4.7e-82	4926	U72459 Chlamydia pneumoniae
gb_pat:AB033806	+ 1885.00	1572.90	1.7e-79	3200	AB033806 Chlamydia pneumoniae
gb_pat:AB1827	+ 1855.00	1571.74	2.0e-79	3200	AB1827 Sequence 1 from Patent
gb_pat:CPN133034	+ 1855.00	1566.44	3.9e-79	6030	AE001311 Chlamydia pneumoniae
gb_pat:AB1849	+ 1811.00	1535.50	2.1e-77	2838	AB1849 Sequence 23 from Patent
gb_pat:CPN133034	+ 1659.50	1400.81	6.6e-70	6110	U59942 Chlamydia pneumoniae
gb_pat:AE001586	+ 1659.50	1400.81	6.6e-70	6234	U59942 Chlamydia pneumoniae
gb_pat:AB033808	+ 1573.00	1322.98	1.4e-65	10574	AE001586 Chlamydia pneumoniae
gb_pat:AB1833	+ 1566.00	1327.35	8.2e-66	2526	AB1833 Sequence 7 from Patent
gb_pat:AB033792	+ 1560.00	1323.96	1.3e-66	2526	AB033792 Chlamydia pneumoniae
gb_pat:AE001587	+ 1442.50	1312.34	2.1e-59	10766	AE001587 Chlamydia pneumoniae
gb_pat:AB1831	+ 1441.50	1312.07	2.2e-59	10044	AE002237 Chlamydia pneumoniae
gb_pat:AB033808	+ 1430.50	1312.73	2.0e-59	3052	AB1831 Sequence 5 from Patent
gb_pat:AE001629	+ 1377.50	1157.36	2.4e-56	10695	AE001629 Chlamydia pneumoniae

seq_name: gb_pat:AB1835

gb_pat:AE002191	- 1377.50	1154.10	3.6e-56	15789	AE002191 Chlamydia pneumoniae
gb_pat:AB033804	- 1148.00	980.22	1.8e-46	1377	AB033804 Chlamydia pneumoniae
gb_pat:AE001360	+ 1120.50	937.41	4.3e-44	14177	AE001360 Chlamydia pneumoniae
gb_pat:AB033796	+ 1089.00	928.16	1.4e-43	1769	AB033796 Chlamydia pneumoniae

seq_documentation_block:

LOCUS	AB1835	2787 bp	DNA	PAT	21-JAN-2000
DEFINITION	Sequence 9 from Patent WO9858953.				
ACCESSION	AB1835				
VERSION	AB1835.1 GI:6731868				
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				

REFERENCE

1 (bases 1 to 2787)

Madsen, A. and Birkelund, S.

NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE

PATENT: WO 9858953-A 30-DEC-1998;

MADSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)

LOCATION/Qualifiers

1..2787

/organism="unidentified"

/db_xref="taxon:32644"

BASE COUNT 811 a 583 c 598 g 795 t

ORIGIN

alignment_scores:

Quality: 4760.00 Length: 928

Ratio: 5.135 Gaps: 0

Percent Similarity: 99.892 Percent Identity: 99.677

alignment_block:

US-09-428-122-2 x AB1835 ..

Align seg 1/1 to: AB1835 from: 1 to: 2787

1	MetLysSerSerPheProlySphValPheSerThrPheAlaIlePhePr	17
1	ATGAAGCTCTTCTTCCCAAGTTGTAATTTCTACATTCTCTATTTCCC	50
17	cleSerMetIleAlaThrGluThrValIleuAspSerSerAlaSerPhe	34
1	TTTGTCTATGATCTCTACCGACAGCTTTGCATTCAAGTCCAGTTTCG	100
34	spGlyAsnLysAsnGlyAsnPheserValArgIuSerGingluAspAla	50
101	ATGGGAAATAAAGATGATATTTTCACTTCGTGAGAGTCAGAAATGCT	150
51	GlyThrThrTrpLeuPheLysGlyAsnValThrLeuGluAsnIleProG	67
151	GGAACCTACCTCTATTAAGGAAATGCTCTGAGAAATATTCCTGG	200
67	YThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAsp	84
201	AACGAGCAGCAGCATCAACAAAGCTCTTTAACACACATCAAGCGGAT	250
84	euthrPheTrpGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla	100
251	TGACTTTCACAGGATCAAGGACCTCTATTTCCAAAGGTGATGCA	300
101	GlyThrValIleGlyAlaValIleAsnSerValValAspLysSerThr	117
301	GGACACTACAGGAGGCTGCTGTTAACAGCAGCGTGTAGATATAATCTAC	350
117	rThrPheIleGlyPheSerLeuSerPheIleAlaSerProGlySerS	134
351	CAGCTTATATGAGGTTTCTTCGCTATCTTTATTCGCTCTCGAAGTT	400
134	erIleThrTrpGlyLysGlyAlaValSerCysSerThrGlySerLeuSer	150

401 CGATACTACCGCAAGAGCCGTTAGCTGCTGCTACGGGTAGCTTGAG 450
 151 LeuThrIysAsnValSerLeuLeuPheSerIysAsnPheSerThrAspAs 167
 451 TTGACAAATAATGTCAGTTGCTCTTCAGCAAAACCTTTTCACGGAAVA 500
 167 nglyGlyAlaIleThrAlaIleThrAlaIleThrSerLeuThrGlyThrMetS 184
 501 TGGGGGTGCTATCACCGCAAAACCTTTCATTAACAGGGACTCAATAGT 550
 184 eAlaLeuPheSerGluAsnThrSerSerIysGlyGlyAlaIleGln 200
 551 CAGCTGCTTTTCTGAAATATCCTCTCAAGAAAGGGGAGCCATTTCAG 600
 201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSe 217
 601 ACTTCGAGTGGCCCTTACCATTAAGTGAACCAAGGAGGAGCTCTTTTTC 650
 217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
 651 TGACAAATACTTCTTCGATCTCGAGTCTGGAGCTGCAATTTTACAGAACCTCGG 700
 234 aLThrIleSerAsnAsnAlaIysValSerPheIleAspAsnIysValThr 250
 701 TGACTATTCTTAATATGCTAAAGTTTCTTATTGACAAATAGATGACA 750
 251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysAl 267
 751 GGACCGAGCTCTCTCAACACCGGGGATATGTCAGAGAGTGTATCTGTGC 800
 267 aTYrLysThrSerThrAspThrIysValThrLeuThrGlyAsnGlnMetL 284
 801 TTATAAACACTAGTACAGATTAAGTACACCTCAGCTGAAATCAGATGT 850
 284 euLeuPheSerAsnAsnThrSerThrThrIleGlyGlyAlaIleTYrVal 300
 851 TACTCTTCAGACAAATACATCGACAACAGGGGAGGAGCTATCTATGTG 900
 301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
 901 AAAAGCTCGAAGCTGCTCCGGAGGACTTACCCTATTCAAGTAGAATAG 950
 317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
 951 TGTCAATGGAGGTACAGCTCTTAAGGTGAGCCATAGCTATCGAAGATA 1000
 334 eArgGlyIleuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
 1001 GTGGGGAATTGAGTTATCCGCCGATAGTGTGACATTTGCTTTTAGGG 1050
 351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
 1051 AATACAGTCACTTCTACTACTCTCGGAGCAATAGAGTAGTAGTGCAT 1100
 367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
 1101 AGGACGAGGCAAGATGACAGCTTGGCTGTGCTGCTGTAGAGCCA 1150
 384 IeTYrPheTYrAspProIleThrThrGlySerSerThrThrValThrAsp 400
 1151 TCTACTTCTTGATCCCATTAAGATCTTCCACAAAGATTTACAGAT 1200
 401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTYrThrGl 417
 1201 GTCTTAAAGTTAATGAGACTCCGGCAGATTTCGCACTACAAATATACAG 1250
 417 yAsnIleIlePheThrGlyGlyLysLeuSerGluThrGluAlaAlaAspS 434
 1251 GAACATCATCTTCACAGAGAAAGTTATCAGACAGAGGCGCAGATT 1300
 434 eLysAsnLeuThrSerIysLeuLeuGlnProValThrLeuSerGlyGly 450
 1301 CTAAATATCTTACTTCGAAAGCTACTACAGGCTGTACTTTCAGGAGGT 1350

451 ThrLeuSerIleuLysHisGlyValThrLeuGlnThrGluAlaPheThrGl 467
 1351 ACTCTATCTTAAACATGAGTGTGAGTCTGACAGTCAAGGATTCACCTCA 1400
 467 ngLysAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGluPro 484
 1401 ACAGGCGAGATTCTCGTCTCGAAATGACAGCTAGGAATCTACTTGAACCTG 1450
 484 IAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
 1451 CTGATACTACACCATTAACAAATTTGTCATTATACATCAGTTCTATAGAC 1500
 501 GlyAlaLysValLysIleGluThrIysAlaIleThrSerIysAsnLeuTh 517
 1501 GGTGCAGAGAGGCAAAATAGAACCAAGCTACGCTCAAAAATCTGAC 1550
 517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTYrGluA 534
 1551 TTATCTGGAACCATCAGTTTATTGAGCCGACGGCAGCTTTTATGAAA 1600
 534 snHisSerLeuArgAsnProGlnSerTYrAspIleLeuGluLeuLysAla 550
 1601 ATCATAGTTTAAAGAAATCCTCAGTCTACGACATCTTAGAGCTCAAAAGCT 1650
 551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyGl 567
 1651 TCTGGAGCTGTAAACAGCACCGAGTACTCCAGATCTATATATGGGTGA 1700
 567 uLysPheHisTYrGlyTYrGlnGlyThrTrpGlyProIleValIrrGlyLT 584
 1701 GAATATCCATTACGGGTATCAGGAACTTGGGGCCCAATGTGTTGGGGA 1750
 584 hGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTYrIle 600
 1751 CAGGGGCTTCTACGACTGCACCTTCACTGAGCTAAACCTGCTATATT 1800
 601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
 1801 CCGAATCCCGAGCGATAGCGCTTATAGCCCTAATAGCTTATGGAATGC 1850
 617 aPheIleAspIleSerSerLeuHisTYrLeuMetGluThrAlaAsnGlnG 634
 1851 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACTGCAAGAG 1900
 634 ILeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhehe 650
 1901 GGTTCAGGAGGAGCCGTGCTTTTGTGTGTGATTAATCACTTCTTC 1950
 651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyGl 667
 1951 CATAGGATAGTACAAAACACAGACGCGGGTTGCCATTGAGTGGGGG 2000
 667 yTYrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
 2001 TTATGTCATAGAGGAAACCTACATACTTGTTCAGATATTAATTTAGTG 2050
 684 IAlaPheCysGlnLeuPheGlyArgAspArgAspTYrPheValAlaLys 700
 2051 CTGCATTTTGTAGCTCTTGGAGAGATAGAGACTATTGTAAGCTAAG 2100
 701 AsnGlnGlyThrValTYrGlyGlyThrLeuTYrTYrGlnHisAsnGluTh 717
 2101 AATCAAGGATAGCTACAGGAGAACCTCTCATTAACGACCAACGAAAC 2150
 717 rTYrIleSerLeuProCysLysLeuArgProCysSerLeuSerTYrValP 734
 2151 CTATATCTCTCTCTTCCAAACTACGCGCTTGTGCTGTTGTTATGTTC 2200
 734 roThrGluIleProValLeuPheSerGlyAsnLeuSerTYrThrHisThr 750
 2201 CTACAGAGATTCTCTCTTTTCAGGAAACCTTAGCTACACCATACG 2250

751	AspaNaApLeuLysThrLysTyrrThrThyProThValIsglyLysE	767
2251	GATACGACTCGAATAACCAGTATACACAACATCTCACTGTTAAAGAAG	2300
767	TTPcLYnaNsApSerPheAlaLeuGlnPheGlyGIYArgAlaProIleC	784
2301	CTGGGGCATGTATATTCCGTTTGCAATTCGGTGGAGAGCCGCATTT	2350
784	YSLeuASgLSerAlaLeuPheGlnLInrMetProPheMetLysLeu	800
2351	GCTTAGATGAAGAAGTCTCTATTATGAGCAATCATGCCCTTCAGAAATTG	2400
801	GlnPheValTYrAlaHisGlnGluGlyPheLysGluGlnGlyThrcGUaI	817
2401	CAGTTTGTCTATGCACATCAGGAAGGTTTTAAAGAACAGGAACAGAAAC	2450
817	aArgGLUpheGLysSerSeraIgleuVaIasNLeuAlaLeuProIIeGlyT	834
2451	TGCTGAATTTGGAAAGTAGCCGCTTGTAATCTTGCCCTTACCTATCGGGA	2500
834	IeaArgPheASPlysGLUserASPcysGlnSPaLatHryrPaenLeuthr	850
2501	TCCGATTGTATAGGATCAGACCTGCCAATATCACACGTACATCTAACT	2550
851	LeuGlyTYrThrValAspLeuValAIrgSerAsnProASPcysThrThrTh	867
2551	CTTGCTTTACTGTGATCTTGTTCTGTAGTAACCCGACACTGTACGACAC	2600
867	rLeuArGlIeSerGLysPseTrPlystrPhelGlyThrAsnLeuAlaIa	884
2601	ACTGGGAATTAGCGGTATCTTGGAATAACCTTCGGTACGAATTTGGCAA	2650
884	rGlnAlaLeuValLeuAIrgAlaGlyAsnHisPhecysPheAsnSerAsn	900
2651	GACAGCGCTTAGTCTCTCGTGCAGGAACCATTTTTCGTTAACTCAAT	2700
901	PheGluAlaPheSerGlnPheSerPheGluLeuAIrgLysSerSeraArgas	917
2701	TTTGAAGCCTTtagccCAATTTCTTTTGAATTCGGTGGTCATCTCGNA	2750
917	nTYrAsnValAspLeuGlyAlaLysTYrGlnPhe 928	
2751	TTACATCTTAGACTTAGAGCAAAATATCCATTC 2784	
<hr/>		
seq_name : gb_bai.AE002235		
<hr/>		
seq_documentation_block:		
LOCUS	AE002235	12127 bp DNA BCT 30-MAY-2000
DEFINITION	Chlamydophila pneumoniae AR39, section 63 of 34 of the complete genome	
ACCESSION	AE002235 AE002161	
VERSION	AE002235.2 GI:8163495	
KEYWORDS	Chlamydophila pneumoniae AR39.	
SOURCE ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
REFERENCE AUTHORS	1 (bases 1 to 12127). Read,T.D., Brunning,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39 Molecular Acids Res. 28 (6), 1397-1406 (2000) 20150255	
JOURNAL MEDLINE REFERENCE	2 (bases 1 to 12127) Read,T.D., Brunning,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Direct Submission	
AUTHORS		
TITLE		
FILE		

JOURNAL	COMMENT	FEATURES
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jun 1, 2000 this sequence version replaced gi:7189672.		
Location/Qualifiers		
gene	source	1..12127
/organism="Chlamydomonas reinhardtii"		
/strain="AR39"		
/db_xref="taxon:115711"		
/note="synonym: Chlamydia pneumoniae AR39"		
complement(150..2987)		
/gene="CP0757"		
/note="This region contains an authentic point mutation, causing a premature stop, and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family authentic point mutation"		
complement(3003..3191)		
/gene="CP0758"		
complement(3003..3191)		
/gene="CP0758"		
/note="hypothetical protein; identified by Glimmer2; putative"		
/codon_start=1		
/transl_table=1		
/product="hypothetical protein"		
/protein_id="AAE38560.1"		
/db_xref="GI:7189673"		
/translation="MRLLSCPMFLIACASFFQGFQEGMGRNQTOSLDANASSIGELF SISKVYSCLEHREINR"		
complement(3336..6086)		
/gene="CP0759"		
/note="This region contains a gene with one or more premature stops or frameshifts, and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family, degenerate"		
complement(6325..9110)		
/gene="CP0760"		
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family, authentic frameshift"		
complement(9393..11918)		
/gene="CP0761"		
complement(9393..11918)		
/gene="CP0761"		
/note="similar to GP:4376270; identified by sequence similarity; putative"		
/codon_start=1		
/transl_table=1		
/product="polymorphic membrane protein G family"		
/protein_id="AAE38561.1"		
/db_xref="GI:7189674"		
/translation="MKRLPFLILSLVPTLSMSNLGAATTEELSAASFDTSTTS FSKRTSATDCTNVPRKDSVIEVPVETGETOSTSCRNKAAGDLNFKGSGSPFS NIDTWTASGAISGEANKTVLTSGEFALSFLKSPASTVTGGALIVNKNSLSDND KVLVDNFSTGDGCALINCAGSLKIANNKSLSIFIGNSSITGGALHTNKLILSSGETL FOGGTAPTAAGKGCAIALADSGTLISIGSDGDIIDEENTIGAGTGVAHSAILDGTSAK IIDLARAOGHTLIYEEDITVGTGSVDALININSPODNKEVGTIVEGEEKTEAEI AKDEKNRKTLLONVARFNKGVIVLKGVGLVANINFGSDANSKLIMDGTSLVANTEI EUNLETINIDISRNGKRITKLSAATAODGITDIPRVLAISDESFOYNGPLNEHSYNG ILIEDAKGDIVIASDRSIDAVOSPITYOGRKWTINSTDNRKATVSNAKGSFRNTAO EAPLPWLVGSLFDVRSIQNFLIELGEGAYPEERFWNAISNVLHSGENORKEFH VSIGAVVAGSTRMGPGDTLSLGFAQLGPARKDFMNPNKTYAGSLRHODHSLYSV VSIILGCGRELILPVYSTLPCSEFYGLSYGTHDRMKTESLPPPLISTDHTSM GGYYWAEEGLTRAVENTSGRFQEOETPEPKVOAYVARODSFVEIGAISRDFSIL YNALIAPIGRIKLERFAOYHYVAAWSYPDCRSRPKCTLLILNSOGSWTKRGSILANO AGIYOASGRFSLCAAALFEQNFGEFMWGSSRSRYNDVAGSIKK"		
BASE COUNT	3532 a	2722 c 2431 g 3441 t 1 others
ORIGIN		

alignment_scores: Length: 928
 Quality: 4746.00
 Ratio: 5.114
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-428-122-2 x AE002235/rev ..

Align seg 1/1 to reverse of: AE002235 from: 1 to: 12127

```
1 MetysSerSerPheProlyPheValPheSerThrPheAlaIlePhePr 17
|||||
9110 ATGAGAGCTCTTTCCCAAGTTTGATTTCTACATTGCTATTCTTCC 9061
17 OleuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPhe 34
|||||
9060 TTGCTGCTATGATGCTACCGAGACAGTTTGGATTCCAAGTCGAGATTCCG 9011
34 spGlyAsnLysAsnGlyAsnPheserValArgGluSerGlnLysAspAla 50
|||||
9010 ATGGGAATAAATAATGGTAATTTTCAGTTCTGAGAGTCAGAGAAGATGCT 8961
51 GlyThrThrTyrlleuPheLysGlyAsnValThrLeuGluAsnIleProG 67
|||||
8960 GGAAGTACCTACCTATTAAAGGAATGTCACCTAGAAAATATTCTCCG 8911
67 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
|||||
8910 AACAGGCAACACCAATCACAAAAGCTGTTTACACACCTAAGGCGCATT 8861
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGluThrValAspAla 100
|||||
8860 TGACTTTTCACAGGTACGGGAACCTCTATTGTTCCAAAACGGTGATGCA 8811
101 GlyThrValAlaGlyAlaIleValAsnSerSerValValAspLysSerTh 117
|||||
8810 GGGAGCTAGACAGGGGCTGCTGTTTAACGACGCGTGATGAATAATCTAC 8761
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
|||||
8760 CACGTTAATAGGTTTCTTCTGCTATCTTTATTGCGCTCTCTGAGAGTT 8711
134 erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
|||||
8710 CGAATACCTACCGGCAAGAGACCTGTAGCTGCTACGGGAGCTTAGT 8661
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheserThrAspAs 167
|||||
8660 TTGCAC.AAAAATGTCAGTTTCTCTTACGCAAAAACCTTTTCAACGGATAA 8612
167 nglyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
|||||
8611 TGGGCGTGTCTATCACCGCAAAAACCTTTTCATTACAGGAGCTTCATGT 8562
184 erAlaLeuPheSerGluAsnThrSerSerLysLysGlyGlyAlaIleGln 200
|||||
8561 CAGCTGCTGTTTTCGAAATATCCTCTCAAGAAAGGCGGAGCCATTCCAG 8512
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyLysValSerPheSe 217
|||||
8511 ACTTCGATGCCCTTACCATTTACGGAACCAAGGGAAGTCTCTTTTTC 8462
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
|||||
8461 TGACAATATCTTCTCGGATTTCTGGAGCTGCAATTTTACAGAAGCCGCG 8412
234 alPhrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
|||||
8411 TGACTATTCTTAATATGCTAAAGTTCCCTTTATATGCAATAAGGCTACA 8362
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysAl 267
|||||
8361 GGAAGCGAGCTCTCAACACAGGGGATATGTCAGAGAGTCTATCTGTGC 8312
```

```
267 aTyrlYsThrSerThrAspThrLysValThrLeuThrGlyAsnGluMetL 284
|||||
8311 TTATAAACTAGTACAGATACATAGGTCACCTCACCCTGGAATACAGATGT 8262
284 euleuPheSerAsnAsnThrSerThrThrAlaGlyGlyAlaIleTyrlVal 300
|||||
8261 TACTCTTCACACACATATACATCGACACAGCGGGAGAGCTATCTATGCG 8212
301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
|||||
8211 AAAAGCTCGAACCTGCGCTCCGAGAGACTTACCTTATTCAGTAGAATAG 8162
317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
|||||
8161 TGTCAATGGAGGTACAGCTCTTAAGGTGGAGGCCATAGCTATCGAAGAA 8112
334 erGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
8111 GTGGGGAATGAGTTTATCCCGCGATAGTGTACATTGCTTTTAAAGG 8062
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
|||||
8061 AATACAGTCACTTCTACTACTCTCGGAGCAATAGAGTATGATCGACTT 8012
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaGlyArgAla 384
|||||
8011 AGGACAGAGCGCAAGATGACAGCTTGGCTTCTGCTGCTGATAGACCA 7962
384 leTyrlPheTyrlAspProIleThrThrGlySerSerThrThrValThra 400
|||||
7961 TCTACTTCTTATGATCCCATTAACATACAGATCATCCACACAGTACAGAT 7912
401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyThrCl 417
|||||
7911 GTCTTAAAGTTAATAGAGACTCGCGAGATTCTGCACTCAATATACAG 7862
417 yAsnIleIlePheThrGlyLysLeuSerGluThrGluAlaAlaAspS 434
|||||
7861 GAAATCATCTTACACAGAGAAAGTTATCAGACAGAGGCCCGCAGATT 7812
434 erLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
|||||
7811 CTAATAATCTTACTTGAAGCTACTACAGCTGTAACTCTTTCAGAGAGT 7762
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrCl 467
|||||
7761 ACTCTATCTTTAAACATGAGAGTGTGACACTGACAGCTCAGGCAATCACTCA 7712
467 nglnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
|||||
7711 ACAGCGAGATTTCGCTCGAATAGGAGCTAGGAACCTACTAGAACCGTG 7662
484 laAspThrSerThrIleAsnAsnLeuValIleAsnLysSerIleAsp 500
|||||
7661 CTGATACGTAGCACCATTAACAAATTGGTCATTACATCAGTTCTATAGAC 7612
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
|||||
7611 GGTGCAAAAGCAAGCAAAAATATAGAAACCAAGCTACGTCAAAAAAATCTGAC 7562
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrlGlu 534
|||||
7561 TTTATCTGAAACCATCATCTTTATGTGACCCGAGCGGACGTTTATGAAA 7512
534 snHisSerLeuArgAsnProGlnSerTyrlAspIleLeuGluLeuLysAla 550
|||||
7511 ATCATAGTTTAAAGAAATCCTCAGCTCTAGACATCTTAGAGCTCAAAAGCT 7462
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyL 567
|||||
7461 TCTGGAAGTGTAAACAAGCACCGCAGTGTGATCTCATTAATAGGTTGA 7412
```

567 uyspshstlyrglytyrgllythrttptglytproilevalttptglyt 584
 7411 GAAATTCATTCCGCTACGCTACGGAACCTGGGCCCAATTTCTTGGGGA 7362
 584 hrglAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrIle 600
 7361 CAGGGGCTTCTACGACTGCACCTTCACTGCACCTTAACTGGCTTATTT 7312
 601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
 7311 CCTAATCCGACGATACGGCTCTTATGCTTATAGCTTATAGCTTATGAAATGC 7262
 617 abhelIleAspIleSerSerLeuHisTyrIleuMetGluThrAlaAsnGluG 634
 7261 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACTGCACAAAGAG 7212
 634 lyleuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
 7211 GGTTCAGAGGACACCGCTCTTTGGTGCTGCATTAATCTTACTTCTTC 7162
 651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
 7161 CATAGGATAGTACAAACACGACGGGCTTCCGCAATTTGAGTGGCGG 7112
 667 YTYrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
 7111 TTATCTCATAGGAGGAAACCTACATCTGTCAGATTAAGATTCTTAGTG 7062
 684 lAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
 7061 CTGCATTTGTCAGCTCTTGGAGAGATAGAGACTATTGTTAGCTAG 7012
 701 AsnGlnGlyThrValTyrGlyTyrThrLeuTyrTyrGlnHisAsnGluTh 717
 7011 AATCAGAGGAGCTCTAGGAGAACTCTATTTACACACACAGAAAC 6962
 717 rTyrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
 6961 CATATCTCTCTTCTTGCAAACTACGGCTTGTGCTTATGTTTC 6912
 734 rothGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
 6911 CTACAGAGATTCCTGCTTCTTTCAGGAAACCTTACGACCATACG 6862
 751 AspAsnAspLeuLysThrLysTyrThrTyrProthValLysGlySe 767
 6861 GATACGATCTAAAAACCAAGATATACACATATCTTACTGTTAAAGAG 6812
 767 rTptGlyAsnAspSerPheAlaLeuGluPheGlyLysArgAlaProIleC 784
 6811 CTGGGGGAAATGATAGTTTCGTTTGAATTCGGTGAAGAGCTCGATTT 6762
 784 yslLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
 6761 GCTTGATGAAAGTCTCTATTGAGCACTATCATCCCTCATGAAATTG 6712
 801 GlnPheValTyrAlaHisGlnGlnGlyPheLysGlnGlnGlyThrGluAl 817
 6711 CAGTTTGCTATGACATCAGAGAGTTTAAAGACAGGAGCAAGAAC 6662
 817 aArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
 6661 TCGTAATTTGGAAGTAGCCGCTTGTGAATCTTGCTTACTTATCGGGA 6612
 834 leaTrpPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
 6611 TCCGATTTGATAGGAATCAGACTCCAGAGATGCAACGACATCTAATCT 6562
 851 leuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
 6561 CTGGTTATACTGCTGATCTTCTTGTAGTAAACCCGAGTACGACACAAC 6512
 867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884

6511 ACTCGAATTCAGCGCTGATTTCTTGAACAACTTTCGTAACGAATTTGCCAA 6462
 884 rGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
 6461 GACAAGCTTTAGTCTCTTGCATGAGGAAACCATTTTGTCTTAACTCAAT 6412
 901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
 6411 TTGGAAGCTTTAGCCATTTCTTTGAATTCGCTGGGTCATCTCGCA 6362
 917 nTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
 6361 TTCAATGTAGACTTAGAGCAAAATACCAATTC 6328
 seq_name: gb_ba1:AE001587
 seq_documentation_block:
 LOCUS AE001587.16448 bp DNA BCT 08-MAR-1999
 DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.
 ACCESSION AE001587 AE001563
 VERSION AE001587.1 GI:4376271
 KEYWORDS
 SOURCE Chlamydia pneumoniae.
 ORGANISM Chlamydia pneumoniae.
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE
 1 (bases 1 to 16448)
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R.W. and Stephens, R.S.
 Comparative Genomes of Chlamydia pneumoniae and C. trachomatis
 Unpublished
 2 (bases 1 to 16448)
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R.W. and Stephens, R.S.
 Direct Submissions
 Submitted (01-DEC-1998) Program in Infectious Diseases, University
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
 TITLE JOURNAL
 JOURNAL
 FEATURES
 source
 1. 16448
 /organism="Chlamydia pneumoniae"
 /strain="GWL029"
 /db_xref="taxon:83558"
 172. 2957
 /note="Polymorphic Outer Membrane Protein G Family;
 contains frame shift"
 172. 705
 /gene="pmp_3.1"
 /note="Cpn0014"
 615. 2957
 /gene="pmp_3.2"
 /note="Cpn0015"
 3196. 4971
 /gene="pmp_4.1"
 /note="Cpn0016"
 3196. 5953
 /note="Polymorphic Outer Membrane Protein G Family;
 contains frame shift"
 4874. 5953
 /gene="pmp_4.2"
 /note="Cpn0017"
 6302. 7786
 /gene="pmp_5.1"
 /note="Cpn0018"
 6302. 9139
 /note="Polymorphic Outer Membrane Protein G Family;
 contains frame shift"
 7787. 9139
 /gene="pmp_5.2"
 /note="Cpn0019"
 complement(9380..11467)
 /gene="Cpn0020"
 complement(9380..11467)
 /gene="Cpn0020"
 misc_feature
 gene
 gene
 gene
 CDS

```

/codon_start=1
/transl_table=11
/product="predicted OMP [leader peptide: outer membrane]"
/protein_id="AADI8172.1"
/db_xref="GI:4376272"
/translation="MKRCFLFASFLVLMGSSADALTTHQEAVKKKNSYLHRSKSGIV
TIEDVLNTHNNLRIOANKVYENTVGSLKLVAGNVMVNRKATLVCDELVEYEDT
DSCILNGRFAMYPWELGSGMTLTPETIVRKYSTSEGKDKLCSGLGYESSD
SLISTGKTLRYCRLPIELFPESIPMEIPKRPINFRGGTGGFGLGYSGYSISR
KHSSTFLPDSFFKRGVGMGFLHCSQKOVPENNMKSAYAHRLAIDMAEADHRYL
HDFCFTHHNVFSGEYHLSDSWETVADTFENNELKNGPTVNDCTNNDNFEYTL
SSVKNVSNFQNAQELPYLTLRQYPISTNGVLENIIVECGFLNFAESDHLVGENFS
LRLAARPKLHKTVPPIPIGLSTLGSLLTYGVDPEISRSQSLSAKLOLDVPLHKK
SYORRHITIEPVTITEERPLAKNEDHIFSIODAFHSLNLKAGIDTSYKTNPR
PRHAKMTITHTLSTESKPEPTACELSPCKKNVSLDAEMIKKCHQDMNI
KREWINGDNVAMTLESLSRSKSLKCDRENITLDVSRPIDOGLDPLSDHNLTKG
LEFRPHPCWNYRLSLRYGWHRODTPNYLETQMLGTKLFEEHQLIGVIEREADSRF
FELKDKPKKPPF"
complement(11484..13190)
/gene="Cp00021"
complement(11484..13190)
/gene="Cp00021"
/codon_start=1
/transl_table=11
/product="predicted OMP [leader peptide]"
/protein_id="AADI8174.1"
/db_xref="GI:4376273"
/translation="MGLPHLTLPGLLGLSLPISTIYAKPPESYGHKLIYISNTOQAL
ATYLEADVIGHDFPVLRKIGEDYLKOSIHSDPQTRKSTIIGAGLAGSSALDVL
QAMETADPLQQLLVSAVSGHLGTSDDLFLKALSPYVILKNAIRLANKTKVI
DHLSFIRKLPEIQLCSAIFLRLETESDAYIDDLAKKSAIRSAIVALOIGYQ
KRFLPTLRMLTSSASPODEAILVALGKLKDSQSYNIKKQKQADVDTLAAOIGY
ALGKEDBALPVIKKQALEERPRALVALRHPSEIGLPIPLFLTKNSEALNVALA
LLEIGDMPKLEIVITERIVOPHNETLALSFGKGTLOMKRVNIIVPODEOEREL
ISTTGLEROIITFLPRLEKFAVLKPCYKLLASOKTQALATISLSHSHOEADIL
FOAAKLPGEPIITRAVDLAIVULTKDPKRSRHLDAKRLIGETLLPDTETQRHPS
MPYLRQYVESRTKLMIDLLETLATIKSESDIRLLQIMTEGDAKNFPVLAGLLIKI
VE"
complement(13172..13762)
/gene="maf"
/translation="Cp00022"
complement(13172..13762)
/gene="maf"
/codon_start=1
/transl_table=11
/product="Maf-type protein"
/protein_id="AADI8175.1"
/db_xref="GI:4376274"
/translation="MSLPVLGSSSPRRKFLIEKRPVPTVIPSNEDESYSYSGDPI
AYTOELAAQKAVASSELHSPDCIILTGDIVSYDGRITFKPODADAIOMLKTLRN
QTHDVSIAVLHKGKLLGSETSOISLTMIPOHRIEYIDVTGLNNGCAYDVCHGGL
ILKRYGCVYVNOGLPIQTLKYLLEELNIDLMDYSL"
complement(13791..15383)
/gene="Yj1K"
/translation="Cp00023"
complement(13791..15383)
/gene="Yj1K"
/codon_start=1
/transl_table=11
/product="ABC Transporter Protein ATPase"
/protein_id="AADI8176.1"
/db_xref="GI:4376275"
/translation="MSIVLDKIGKSLGTRILFPDDVSVPENPGNCYGLTGNAGKSTL
LKITMIEPTRGISILPKKVGILRONDISFHDITVLDGVINGNTRIMFALORPNLY
LQFPAIGMELEIEITIGEENGRADSEAEELGKTGEGIVIVVSHRHLNIDIT
RYVLCOALFGHEPALLDEPTNHLIDYLSINWLGNFRLGDIIVVSDKRAKMAPIIDIT
HIADIDYDTIIVPGNYDDMEKMTASROEKADISKSKKLSLEIKFEYAKGASRA
SOVSRLEIKTPOELKKSNIORPYIRFPLSDKSGGVLSLEIKFEYAKGASRA
PSELTIOGDKLIGINNGIKGTTLMKLAGVAPSSGSIKIGHOALISYFQONSDY
LADCGEITLFEWLRNKTSTINDOETRSYVGLKALFGSDDAFKOIALSGEETARLLMAG
MMLNHNVLILDEANNHLDLESYALSMAINDKGTATFVSHDRGLIQDCATRLIFD
KXKITFFDGTMDYTAGHROL"
complement(15438..16376)

```

```

CDS
/gene="xerc"
/translation="Cp00024"
complement(15438..16376)
/gene="xerc"
/codon_start=1
/transl_table=11
/product="Integrase/recombinase"
/protein_id="AADI8177.1"
/db_xref="GI:4376276"
/translation="MIAISYSPLDYILKMKVASPPTLRNYCLDNLGKLFLEERGNLA
PSSLOATEKRYSELPSLFTENHRMYIAKIEGKAKRITKRLSSIKSPAHYC
VQIQLNENREITHGRLKRELPSPTVQVRYVATNPISYHGRIDRCILMELYS
SGLRISLIVANAKQDEFLSTHLIRIKRGKERRIIPVTSNAIQIQLYLNHDKRLK
KDPQAFILNFRGRISTRSIDRSQEYLRSGLSGLITPHTIRRTIATHWLESMDLK
TIOALHSSLETTTYYTQVSVKLRQTHQEAHPHA"
BASE COUNT 4710 a 3330 c 3665 g 4743 t
ORIGIN

alignment_scores:
Quality: 4746.00 Length: 928
Ratio: 5.114 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-428-122-2 x AE001587 ..
Align seg 1/1 to: AE001587 from: 1 to: 16448

1 MetLysSerSerPheProLysPheValAlPheSerThrPheAlaIlePhePr 17
|||||
172 ATGAAGCTCTCTTCCCAAGTTGTATTCTTCTACATTGGATATTTCC 221
17 cLeuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPhea 34
|||||
222 TTGTCTATCATTTGCTACCGACAGCATTTTGGATTCAATGGCAGTTGG 271
222 spLysAluLysAsnGlyAsnPheSerValArgGluSerGlnLysAla 50
|||||
272 ATGGGATATAAATGGTAATTTTTCAGTTCGTGAGCAGACGAAATGCT 321
272 GlyThrThrThrLeuPheLysGlyAsnValThrLeuGlnAsnLeuProL 67
|||||
322 GGAACCTACCTACCTATTTAAGGGAATGTCACCTGAGAAATATTCTCG 371
322 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
|||||
372 AACAGCAGCAGCAATCAGAAAAGCTGTTTAAACACACTAAGCGGATT 421
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
|||||
422 TGACTTCAAGAGTAAGGGAACCTCTATTGTTCCAAACGGTGATGCA 471
101 GlyThrValAlaGlyAlaValAsnSerSerValValAspLysSerThr 117
|||||
472 GGAACCTAGCAGGCGCTGCTTAAACAGCAGCGTGATGATTAACACTAC 521
472 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySerS 134
|||||
522 CAGCTTATAGGGTTTCTTCGCTATCTTTATTCGCTCTCCGGAAGTT 571
134 eutLeuThrThrGlyLysGlyAlaValAsnSerSerThrGlySerLeuSer 150
|||||
572 CGATAACTACCGGCAAGAGCGCTTAGCGCTACGCGGTAACCTGAGT 621
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAs 167
|||||
622 TTGAC AAAAATGTCAGTTTGCCTTCAGCAAAAACCTTTTCAACGATGA 670
167 nglyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMetS 184
|||||
671 TGGCGGTGCTATCAGCGCAAAAACCTTTTCATTTAACAGGACTACATGT 720

```



```
184 eRAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
|||||
721 CAGCGCTGTTTCTGAAAAATACCTCTCAAGAAAGGCGGAGCCATTGAG 770
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValIleSerPheSe 217
|||||
771 ACTTCGCAATGCCCTTACCATTAAGTCTGAAACCAAGGGAAGTCTTTTTC 820
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
|||||
821 TGACAATATCTTCTCGAATTCGAGCTGCAATTTTACAGAACCTCGG 870
234 aThrIleSerAsnAsnAlaLysValIleSerPheIleAspAsnLysValThr 250
|||||
871 TGACTATTCTTAATATAGCTAAAGTTCTTTATTGACAAATAGGTCACA 920
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleGlyAsp 267
|||||
921 GGAGGAGCTCTCTCAACAACGGGGGATATGTCAGAGAGTGCATCTGTC 970
267 aTyrlLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
|||||
971 TTATTAACACTAGTACAGATTAAGTCACTCACTGGAATCAGATGT 1020
284 euLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTyrlVal 300
|||||
1021 TACTCTTAGCAACAATATCATCGACACAGCGGAGGAGCTATCTATGTG 1070
301 LysLysLeuGlnLeuAlaSerGlyLeuThrLeuPheSerArgAsnSe 317
|||||
1071 AAAAAGCTCAAGCTGCTCCGAGAGCTTACCTATTACAGTAGAATAG 1120
317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
|||||
1121 TGTCATAGGATGACAGCTCTCTAAAGGTGAGCCCATGCTATCGAAGTA 1170
334 eRgLyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
1171 GTGGGAAATGAGTTATCCCGCCTAGTGGACATGTCTTTTACG 1220
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
|||||
1221 AATACAGCTACTTCTACTACTCTCTGGAGCAATAGATAGTATGACTT 1270
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaGlyAlaI 384
|||||
1271 AGGAACGAGTGCAAAGATGACAGCTTTCGTTCTGCTGCTGATAGAGCA 1320
384 leTyrlPheTyrlAspProIleThrThrGlySerSerThrThrValThrAsp 400
|||||
1321 TCTACTTCTATGATCCCATTAAGTACAGATCATCCAAACAGTTACAGAT 1370
401 ValLeuLysValAsnGlnThrProAlaAspSerAlaLeuGlnTyrlThrGl 417
|||||
1371 GTCCTTAAGGTAAAGAGACTCCGAGATTTCTGCATACATATATACAGG 1420
417 yAsnIleIlePheThrGlyGluLysLeuSerGlnThrGluAlaIleAspS 434
|||||
1421 GAACATCTCTTCCACAGAGAAAAGTTATCAGACAGACAGAGCGCGAGATT 1470
434 eRlysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
|||||
1471 CTAATAATCTTACTCGAAGCTACACAGCCGTAACTTTACAGAGGT 1520
451 ThrLeuSerLeuLysHisGlyAlaThrLeuGlnThrGlnAlaPheThrGl 467
|||||
1521 ACTCTATCTTAAACATGAGAGTACTCTGCACTCAGACTCAGCAATTCACA 1570
467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGlnProA 484
|||||
1571 ACAGGCAATTTCTCTCGAAATGAGACTAGAACTACTAGAACCTG 1620
484 laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
|||||
1621 CTGATTAAGACACCAATAATTTGGTCAATTAACATCAGTCTATAGAC 1670
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
|||||
1671 GGTCAAAAGAGCAAAATAGAAACCAACACTACGTCAAAATAATCTGAC 1720
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrlGlu 534
|||||
1721 TTTATCTGGAACCATCACTTATTTGAGACCCGACGCGACGTTTATGAAA 1770
534 snHisSerLeuArgAsnProGlnSerTyrlAspIleLeuGluLeuAla 550
|||||
1771 ATCATAGTTTAGAAATCTCAGTCTTACAGACATCTTAGAGCTTAAAGCT 1820
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
|||||
1821 TCTGGAAGTAAACAGACCGCAGTGCATCTCAGATCTTATATGGGTGA 1870
567 uLysPheHisTyrlGlyTyrlGlnGlyThrTrpGlyProIleValTrpGly 584
|||||
1871 GAAATTCATTACGCTATCAGGGAAGTCTGGGGCCCAATTGTTGGGGGA 1920
584 hrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrlLe 600
|||||
1921 CAGGGGCTTCTACAGACTGCAACCTTCAACTGACTTAAACTGCGCTATATT 1970
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
1971 CCTAATCCCGAGCTATCGGCTTGTAGTCCCTAATAGCTTATGGAATGC 2020
617 aPheIleAspIleSerSerLeuHisTyrlLeuMetGluThrAlaAsnGlu 634
|||||
2021 ATTTATGATATTAAGCTCTCTCCATTTATCTTATGAGACTGCAACAGAG 2070
634 lYleuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
|||||
2071 GGTTCAGAGGAGACGCTGCTTTGGTGGTGTGATTAATCACTTCTTTC 2120
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyGl 667
|||||
2121 CATAGAGTATGATCAAAAACACAGCGGGGTTTCGCCATTGTAAGTGGCG 2170
667 yTyrlValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
|||||
2171 TTATGTCTATAGAGAAACCTACATCTTGTTCAGATTAAGATTCTTATGTG 2220
684 laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrlPheValAlaLys 700
|||||
2221 CTGCATTTTTCACAGCTCTTTCGAGAGATAGAGACTTGTGTAGCTAAG 2270
701 AsnGlnGlyThrValTyrlGlyGlyThrLeuTyrlTyrlGlnHisAsnGluTh 717
|||||
2271 AATCAAGGTACAGTCTACGAGGAAGCTCTATTATCAAGCAACAAGAAC 2320
717 rTyrlIleSerLeuProCysLysLeuArgProCysSerSerLeuSerTyrlAlP 734
|||||
2321 CTATATCTCTCTTCTTGCACAACTACGCGCTTGTGTTGTTATATGTTTC 2370
734 roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrlHisThr 750
|||||
2371 CTACAGAGATTCCTGTTCTCTTTCAGGAAACCTTACTACACCCATACG 2420
751 AspAsnAspLeuLysThrLysTyrlThrThrTyrlProThrValLysGlySe 767
|||||
2421 GATAACGATCTGAAGAACCAAGTATACAAATATCTCTAGTTAAAGGAAG 2470
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyGlyArgAlaProIleC 784
|||||
2471 CTGGGGGATATAGTTTGGTTAGAAATTCGTTGGAAGAGCTCCGATTT 2520
784 ySLeuAspGluSerAlaLeuPheGluGlnTyrlMetProPheMetLysLeu 800
|||||
```

```

2521 GCTTAGATGAAGTCTCTATTGAGCATACATGCCCTTCATGAATTG 2570
801 GlnPheValIYrAlaHisGIngluGlyPheGluGlnIYthrGluAl 817
2571 CAGTTTGTCTATGCACATCAGGAAGGTTTAAAGAACAGGGAACAGAGC 2620
817 aaAGGluPheGlySerSerArgLeuValaLanLeuAlaLeuProIleGlyI 834
2621 TCGTGAATTTGGAGAGTACCGCTCTTGATCTTGCTTGCCTTACCATCGGA 2670
834 leaArgPheapLysGluSerAspGlyGlnaspAlaThrIYrAsnLeuThr 850
2671 TCCGATTTGATAGGAATCAGATGCCAAGATGCAAGCTACAACTACT 2720
851 LeuGlyYrThrValaLapLeuValaIYrSerAspProspGlyThrThr 867
2721 CTTGGTTATCTGCTGATCTTGCTGATGACCCGACGTGTACGACAC 2770
867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884
2771 ACTGCGAATTAAGCGGTGATCTTGAAACCTTCGTACGATTTGGCAA 2820
884 rGglAlaLeuValaLeuArgAlaGlyAsnHisPheGlySerAsnSerAsn 900
2821 GACAACTTATAGTCTCTGTCGACAGGACCATTTTCTTAACTCAAT 2870
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgIYrGlySerSerArg 917
2871 TTGGAACCTTACCCAAATTTCTTTGAAATTCGTGGGTCATCTCCAA 2920

seq_name: gb_ba2:CPN133035

seq_documentation_block:
LOCUS CPN133035 26920 bp DNA BCT 24-FEB-1999
DEFINITION Chlamydia pneumoniae GGA1-protein family cluster B, genes omp6,
            hb1, hb2, hb3, hb4, hb5, hb6, hb7, omp7, pseudogenes omp8, omp9,
            omp15, strain VR1310.
ACCESSION AJ133035
VERSION Hb1 gene; HB2 gene; hb3 gene; hb4 gene; hb5 gene; hb6 gene; hb7
KEYWORDS gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer
            membrane protein 6; outer membrane protein 7; pseudogene.
SOURCE Chlamydia pneumoniae.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE Boesen, T.
AUTHORS Submitted (21-JAN-1999) Boesen T., Department of Medical
TITLE Microbiology and Immunology, University of Aarhus, The Bartholin
JOURNAL Building, DK-8000 Aarhus, DENMARK
2 (bases 1 to 26920)
REFERENCE Hjerne, K., Boesen, T., Daugaard, L., Knudsen, K., Madsen, A.,
AUTHORS Christiansen, G. and Birkelund, S.
TITLE Chlamydia proteins containing the GGA1-repeat belong to a subfamily
JOURNAL of auto-transferring pathogenicity factors
FEATURES
    source
        location/Qualifiers
            1..26920
                /organism="Chlamydia pneumoniae"
                /strain="VR1310"
                /db_xref="taxon:83558"
                /gene="omp6"
                /gene="omp6"
                /gene="omp6"
                /note="putative"
                /codon_start=1
                /transl_table=11
    sig_peptide
    CDS

```

```

/product="outer membrane protein 6"
/protein_id="CAB37075.1"
/db_xref="GI:4455891"
/translation="MRESLGPPPLVFSLTLLSVFDSLSATITSLPDESEI
ERSYNVAGGVYSLTGVSISVNDNALNACNVNVSIGVTFPAGNHGLEYFN..sSGT
TRBSAVLCCDDPQATRFSGFTSLFISFGSDIKBOGCLSKNALMLNNYVRFPPON
QSKTKGALISGANVTIVGNVDSVFYONATFEGLAHSSCPLOTIANQAEIRAQVTA
KNGSGALVSDGDIDIDONAVVLFRENEALTTAIGGAVCCCLPTSGSSTPVPVTF
DNKOLVERNRHSINGGAIYARKLSISGGPTLFINNISYANQONLGAIAIDGSEI
SLASEKTTIFEOGNRTSLPFLNGHLHONAKPFLKLOANCGSTIEFDPIFEADGSTO
LININGDPKNEVEFTGTLFSGEKSILANDRPDESTIQNVNLSAGVYIRKGAFTVSK
TOSPSGSHVLDTGKRLIASKEDAIATGLAIDIDSSSTSAVITAMNTANKRQISVTD
STELISTGNAVEDLRLNRNSQTFPLSLFEGGASVTVIAGDLPVSPHFGGNMKT
AMTGTGNKVEEFEMWDKINYPREPKEGNLVPNLNMGAAVNRSLMOVQETRHASLTD
RGLMIDICGNFHFVYASSEDNIRYRHNNGGVYLVNNEITPKRHTSNAPLSQLEFRDQY
AVSNNEYRMYLSLYOYTTSLGNIFFYASRNPNVNGILSRFLQPLMIFHLLAY
GHATNDKMTQYANPMPVKNRNNGCAIECGSPMLPLFENGRLFOGAIPEMKLOLVY
AYQGDPEKTTADGRFSNGSLTISVPLGIRFEKIALSDOVLDFDSFSYIPDIERDP
SCENALYISGDSWLPAAHVSRHAFVSGTGRHFHNDYTELLRCGSLIECPHARNINI
NCGSRFRP"
498..3266
/gene="omp6"
3976..6870
/gene="hb1"
3976..6870
/gene="hb1"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="Hb1 protein"
/protein_id="CAB37076.1"
/db_xref="GI:4455892"
/translation="MRKRLIATVIALSIIILAGVLLTVAIPGLISVITSPAGNG
ACALGCVMALGIDVLLKREVPVILASVTTPTGSPRGISISGADSTIRSLPTL
LDEGHPOSMRKRLIATVIAVIFESIIILASGVLLTVAIPGLISVITSPAGMALG
VMAIGIDVLLKREVPVILASVTTPTGSPRGISISGADSTIRSLPTLPTDEGH
OSMRKRLIATVIAVIFESIIILASGVLLTVAIPGLISVITSPAGMALGCVLALG
IDVLLKREVPVILASVTTPTGSPRGISISGADSTIRSLPTLPTDEGHPOSMRK
VESHLNNKSLPYDGHLEBKTKHQIRVARSLSKAMPEDLIRIRIEEEFEFLSKR
KRLIDATTVIERKILTEOLERNMLRKAFTSYLDSDISFKIKINDFEKLAKFMILNSK
ICRFTIIFENHEHGVAKSILHKNVLLLEKVIYRSLOKSYRIDGMSAKMKILGNPFE
SLDEPKTIKMEHAKMELESLSYRKVFLASDEVVPTSPDKWDSGIKPRALSIE
ISRDQWQRKHLKHQESLYIQARDRLTDOSSKRNQKELKAEQEIYSMEYRKKEI
ERVERIRAIQKLLPNILREBEETGCEYVTPYQGTASSDLTDIIGRIEVSREBN
QNOESCVAKLVNLSHEVMSWEVKQYGRKKKEFDQMSLEFTEHTELELVLDKDY
KHLSTFKVNNKREVOYAKFKLVLESDLEGLIADQESAEILAOEELPLIATRGAL
KAVFGSLCCALASKAKPYEEDPFDODSDIQLRALRLRQEAASLEDEIKRFSNLE
NDIAEERLLKESQTERFAGLVLEIAVSTYDLRSLTMTGEGPESEKVVSMYL
NYNVEKRRAKTRIVEMTQRYRDFKMALEMQFNEALLQELSLQAPSE"
7150..8058
/gene="hb2"
7150..8058
/gene="hb2"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB2 protein"
/protein_id="CAB37077.1"
/db_xref="GI:4455893"
/translation="MOVFLSPQDLPPPRRSLGVSCSKRLSLITLVLLVGLLTLGIP
GLTACISGAGLISGALGVIVISGLIFLIVRVPVYRSEIRIGVSVTPSEPALE
KQKPEPKTILDRPKLELDOLDIYIOGVFCLIRIDPKYEDGGLITLEAKEXLRVPD
VAKQKMSSEFLDIORVLENEAVYVEHCODPLENTAYELFSSQELRDYICAGVCYLP
GDARADRLKRSYKVEVMDRFMRVTWKSWSVASVLDHSYGVALELKKAVGLVESVYKI
LFKSYRADVFECEKAKIORDGRFVWL"
8185..9492
/gene="hb3"
8185..9492
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB3 protein"

```

```
/protein_id="CAB37078.1"
/db_xref="GI:445894"
/translation="MMLDRYADKFLIREKEEKMERHLEFATVYRKASGHAYAKAKAA
FEKERSNENORXKVDVEKWLKGLAEFRNOSRRARERLQTLYPEVSEVERLIER
ORFKVLENTLYADIIEKKYHCVREOEYHKAEYRENGEVLISAEEVSECO
RLDCLETWMSKILTKAESVFEEMKPDATKELGNKVLSDVNRLEIICEDAEEMPIFE
EITEMLRMVELPLPKNTEFEKASLOYSCKEMLAYKVEPOCKESPITYRSOELELN
ODLQATVNCQERLQGSFSDLESKVRTRCDRLRQMKHFEVQGLNFINEELLVGAEFL
TOARLDVATVPMEFYLOYHNIRKREVRQOMAKTEREIRQAFQGVKEDLLAED
TILKEEDYWLDRDMLRDERKNRRLICNKIAAQORVKGK"
9805.10698
/gene="hb4"
9805.10698
/gene="hb4"
9805.10698
/gene="hb4"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB4 protein"
/protein_id="CAB37079.1"
/db_xref="GI:445895"
/translation="MLLISGALFLTGLTIGLSAITSFGIGLSALGGVLMISGLIC
LIVKREPIVPEPIEGVSLAPSEPALQAQKTLAQPKELDQDLDIQEYFACL
KLKDSYERSFNLDAKKELVDFVVEDLSEIFELRQIYAOQMDLNLINGRSL
MTAESSELDLFFHVKRIGYLPBGDVREGGLKSKSIYARLMSLHCIEHKVAVADR
NSYAMAERAKALGALGESVYRSITQSYRDKFLSESRAKIPNQHITMLRDQAKSGC
AEKLGMPRVGKNGKQSRG"
10776.12119
/gene="hb5"
10776.12119
/gene="hb5"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB5 protein"
/protein_id="CAB37080.1"
/db_xref="GI:445896"
/translation="MDEIFPHELYEKTFLKRLHRKQALAKTTEKKRSKKNLOAYEE
ANARLKYVDWYDOEFQKAGERLEKLHALYPRVSVSIRENKIOETRSNKEAYEIE
ENRCVCREQEDWYKKEEKREAFERERGNKILSPDELSEDFDGLKFKSKMEL
EGHILKQKATAEVEKILISDNESREIYFEDVKEKPECEIEIEKTLRAELPLPT
KKAFEKACSOYNSCAEMLEKVPYCKESLAVYSKEKLVSLDDELRRAATECCKRQG
DSGLSEVRACRQLREIRIOEFTOGLVEKELCYSSRLNTEDCDVSQVKEKAP
GKKFYQYDEIYRVRQSRMTMSERLREGVACNMLKAGLSEEDKVLKEEYMLY
REERKNKEKRLVGTFKIVATQRIQEFOPSDIYVSSNKKVSLMDKARFLFRBDS"
12362.12987
/gene="hb6"
12362.12987
/gene="hb6"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB6 protein"
/protein_id="CAB37081.1"
/db_xref="GI:445897"
/translation="MTAAPAILHVSPTPEETKVIIPKDSKSNALGITLIVGILLV
CGAIVSGVSGSLIVCGISITSLIGVFLGLILLRLRELLERLQIEAKOLAE
TFDEKLKELMYOTSEKLEKIEGSRYSNOGFLNATOKIILDESISITSEPDL
RQIFDEKIEFILSGERLEFIANLFRKQGDVYLNGLNADIAIVGPNMYKAMVIE
KAAVAVHEFIVLTLMARELEFF"
13014.14585
/gene="hb7"
13014.14585
/gene="hb7"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB7 protein"
/protein_id="CAB37082.1"
/db_xref="GI:445898"
/translation="MYOENLRLELRLLYNSVOKSVADRLLFSYKTKMVDPTLIPPEE
DKKCAAEKAPLEOKLILDYKSTFWLNEENETIUNDPMSGLNTVTRKRYQOYD
DSERMHKYLQKLEDNTEKLLDESSSESTEANKLLSDLVDLBDAKTLFVAKQOE
VETRVKDLRARVGTVPKQDTEAKRKVELLEASLETFLDSISELVQCLEDDIYWE
```

```
QVVKDLARTQELBEDDIEAKREDAEDLSLNERLKSSTMLDRKWHIENEDSITW
WTSQLEMKDKMARLKLKEDITVSYPEIDEITCISLEELPLTTRRELLTKSYLKFKI
CSETLLKMTSVENNIYVOYEVOQLNLFKQGSORFGKQODFANIEQVALQK
RRELQNEIEIGGFNFMKEDFKAANDLYIRSTABOKNMFVPCHELFRYHEVYK
LLEIMVNCDSYRDKAKKCSLRDEKELQNEIKKEERYOKKQQRHADRSHITTYOK
LIAEELAELELKKT"
14954.17479
/gene="omp7"
14954.17479
/gene="omp7"
/note="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein 7"
/protein_id="CAB37083.1"
/db_xref="GI:445899"
/translation="MKIPRFLILSVPLTSMNSNLGAATTEELASNSFQGTSTTS
FSKTSSTARDGTNYVFKDSVLENNPKTGETQSTSCFKKDAAGDLNFGGFSFTFS
```

```
alignment_scores:
  Quality: 4746.00      Length: 928
  Ratio: 5.114          Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
US-09-428-122-2 x CPN133035 ..
```

```
Align seg 1/1 to: CPN133035 from: 1 to: 26920
```

```
1 MetlyssSerPheProlyPheValPheSerThrPheAlaIlePhePr 17
|||||
17762 ATGAGTCTTCTTCCCAAGTTTGATTTCTCATTTGCTATTTTCCC 17811
17812 TTGTCATGATTCCTGCTACCGAGACAGTTTGATTCAGTGCAGTTTGC 17861
17812 TTTGTCATGATTCCTGCTACCGAGACAGTTTGATTCAGTGCAGTTTGC 17861
34 SPGLYASNLVASNGLYASNPheserValargIuserGlnGluAspAla 50
|||||
17862 ATGGGAATAAAGATGTAATTTTCAGTCTCGTAGAGTCAGAGATGCT 17911
51 GLYThrThrTyLeuPheUlyGlyAsnValThrLeuGluAsnIleProG 67
|||||
17912 GGAACCTACCTACCTATTAAAGGAATGTCTCTAGAAAATATCTCTGG 17961
67 YTHrGLYThraAlaIleThrIlySerCysPheAsnAsnThrIlySclyAsp 84
|||||
17962 AACAGGCACACCAATCACAAAAGCTGTTTAAACACACTAAGGCGATT 18011
84 eutrrPheThrGLYAsnGlyAsnSerLeuPheGlnThrValAspAla 100
|||||
18012 TGACTTTCACAGGTACGGGACTCTCTATTGTTCCAAACCGTGGATGCA 18061
101 GLYThrValAlaGlyAlaValaValAsnSerSerValAlaAspIlySerTh 117
|||||
18062 GGGACTGAGCAGGCGGCTGCTTAACAGCAGCGTGTAGATTAATCTAC 18111
117 rTHrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlyIlys 134
|||||
18112 CACGTTATAGGTTTCTTCCTGCTATCTTTATGCGTCTCCCTGGAAGTT 18161
134 erIleThrThrGLYGLYGLYAlaValaSerCysSerThrGlySerLeuSer 150
|||||
18162 CGATTAACACCGGCAAGAGACCTTAGCTGCTACGCGGTAGCTTAGT 18211
151 LeuThrIlyAsnValSerLeuLeuPheSerIlyAsnPheserThrAspAs 167
|||||
18212 TTGAC AAAAATGTCAGTTGCTCTCTCAGCAAAAACCTTTCAACAGATAA 18260
167 nGlyGLYAlaIleThrAlaIlyThrLeuSerLeuThrGLYThrThrIlys 184
|||||
18261 TGGCGGTGCTATCACCGCAAAAACCTTTTCAATTACAGGACTACATGT 18310
```

```

184  erAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
|||||
18311 CAGCTGTGTTTGTGAAATACCTCTCAAGAGAGCGGAGCCATTGAG 18360
201  ThrSerSpAlaLeuThrIleThrGlyAsnGlnGlyValSerPhe 217
|||||
18361 ACTTCGAGTCCCTTACCATTACGGAACCAAGGGAGTCTTTTTC 18410
217  rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
|||||
18411 TGACAATACCTCTTCGAGATTCTGGAGCTGCAATTTTACAGAACCTTCGG 18460
234  aThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
|||||
18461 TGACTATTTCCTAATATGCTAAAGTTTCCTTATTGACAAATAGGTGCA 18510
251  GlyLaseSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysAl 267
|||||
18511 GGAGCGAGCTCTCAACACAGGGGGATATGTCAGAGGTGCTATCTGTGC 18560
267  aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
|||||
18561 TTATAAACTGTACAGATACAGTACAGTACCCCTCAGCTGGAATCAGATGT 18610
284  euLeuPheSerAsnAsnThrSerThrThrAlaGlyGlyAlaIleTyrVal 300
|||||
18611 TACTCTCAGCAACATATACATCGAACAGCGGGAGGAGCTATCTATGTG 18660
301  LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
|||||
18661 AAAAAGCTCGAGCTGCTCCGGAGGACTTACCTATTCAGTAGAATAAG 18710
317  rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAsp 334
|||||
18711 TGTCATGAGAGTACAGCTCCTTAAGGTGAGCCATAGCTATGCAAGATA 18760
334  ergGlyLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
18761 GTGGGGAATTGAGTTTATCCGCCATAGTGTGACATTGCTTTTATGGG 18810
351  AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspIe 367
|||||
18811 AATACATCTCATTCTACTACTCTCGGAGCAATGAGAGTATGATGACTT 18860
367  uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
|||||
18861 AGGAACGAGTGCAAGATGACAGCTTGGCTTGTGCTGTAGAGCCA 18910
384  leTyrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
|||||
18911 TCTACTCTCTATGATCCCATACCTACAGATCATCCACACAGTTACAGAT 18960
401  ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
|||||
18961 GTCTTAAAGTTATAGACTCCGGCAGATTCTGACTACATATTAACGG 19010
417  yAsnIleIlePheThrGlyLysLysLeuSerGluThrGluAlaIleAsp 434
|||||
19011 GAACATCATCTTCACAGAGAAAAGATTATCAGACAGAGCCGACAGATT 19060
434  erLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
|||||
19061 CTAAAAATCTTACTCGAAGCTTACTACAGCTGTAACTCTTTCGGAGGT 19110
451  ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
|||||
19111 ACTCTATCTTTAAACATGGAGTACTCTGACAGCTCAGGCATTCACTCA 19160
467  nGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGluProA 484
|||||
19161 ACAGGCGAGATTCTGCTCGAAATGACGTAGGAATCTCTAAGAACCTG 19210
484  laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500

```

```

19211 CTGATCTACACCATTAACAATTTGGTCTATTAACATCACTCTATAGAC 19260
501  GlyAlaLysLysAlaLysIleGluThrLysAlaIleThrSerLysAsnLeuTh 517
|||||
19261 GGTCGAAAGAGCGAAAATATAGAAACCAAGCTACGTCAAAATAATCTGAC 19310
517  rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
|||||
19311 TTTATCTGGAAACCATCTCTTATTGGACCCGACGGCAGCTTTTATGAAA 19360
534  snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGlnLeuLysAla 550
|||||
19361 ATCATAGTATTAGAAATCCTCAGCTCAGACATCTTAGAGCTCAAGACT 19410
551  SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
|||||
19411 TCTGGAAGCTGTACAAACACCGCAGTGCAGATCCGATCTAATATGGGTGA 19460
567  uLysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGlyT 584
|||||
19461 GAATTTCCATTACGGCTATACGGGAAGTGGGCCCATTTGTTGGGGA 19510
584  hrgLysLaseThrThrAlaThrPheAsnTrpThrLysThrGlyTyrIle 600
|||||
19511 CAGGGCTTCTACGACGTCGAACCTCAACTGAGACTAAACTGCGCTATATT 19560
601  ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
19561 CCTAATCCGAGCGGTATCGGCTTTTATGCTCTTAATAGCTTATAGGATGC 19610
617  aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGluG 634
|||||
19611 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACTGCAAGAGAG 19660
634  LysLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
|||||
19661 GGTTGCGAGGAGACCGTCTTTTGGTGTGCTGAGTATCTTACTTCTTC 19710
651  HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
|||||
19711 CATAGAGATAGTCAAAACACGACGCGGGTTTCGCCATTGTGAGTGCGG 19760
667  yTyrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
|||||
19761 TTATGTCATGAGGAGAACTACTACTGTGTCAAGTATAGATCTTAGAG 19810
684  laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
|||||
19811 CTGCATTTTGTGAGCTCTTGGAAAGATAGAGACTCTTGTAGGCTAAG 19860
701  AsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGluTh 717
|||||
19861 AATCAAGGTACACTACGAGAGAACTCTATTACAGACACAGCAAGAAC 19910
717  rTyrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
|||||
19911 CTATATCTCTCTCTCTGCAAACTACGGCTTGTGTGTCTTATAGTTC 19960
734  roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
|||||
19961 CTACAGAGATCTCTGTTCTCTTTCAGAAACCTTAGTACACCCCTAAG 20010
751  AspAsnAspLeuLysThrLysTyrThrThrTyrProThrValLysGlySe 767
|||||
20011 GATACAGATCTGAAACCAAGTATACACATATCCACTGTTTAAAGAGAG 20060
767  rTrpGlyAsnAspSerPheAlaLeuGlnPheGlyGlyArgAlaProIleC 784
|||||
20061 CTGGGGGAATGATAGATTTCGCTTAAATTCGATGGAAGAGCTCCGATT 20110
784  ysLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
|||||

```

```
20111 GCTTAGATGAAGTGTCTATTGACGACATGCTCCCTTCATGAAATTG 20160
      801 GlnPheValIYrAlaHisGlnGluGlyPheYusGluGlnGlyThrGluAl 817
      20161 CAGTTGTCTATGACATCAGGAAGCTTTAAAGACAGGAGCAAGAAC 20210
      817 AARGGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
      20211 TCGGAAATTTGGAACTACCGCTCTTGCAATCTCCCTTACCTACCGCA 20260
      834 LeArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
      20261 TCCGATTTGATAGCAATCAGACTGCGCAAGATGCAAGCTACATTAAGT 20310
      851 LeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTr 867
      20311 CTGGGTATACTGTGATCTGTGCTAGTAAACCCGACTGACAGACAAC 20360
      867 rLeuArgLLeSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884
      20361 ACTGGAAATTTAGCGGTGATCTTGGAACAACTTCGTCGGAATTTGGCA 20410
      884 rGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
      20411 GACAAAGCTTTAGTCTCTGTCGACAGGACCAATTTTGTCTTAACCAAT 20460
      901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
      20461 TTGGAAGCCTTTAGCCAAATTTCTTTGAATTCGTCGTCATCTCGCA 20510
      917 nTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
      20511 TTACATATTAACCTTAGAGCAAAATACCAATTC 20544

seq_name: gb_bai:AP002545
seq_documentation_block:
LOCUS AP002545 300050 bp DNA BCT 07-JUL-2000
DEFINITION Chlamydia pneumoniae genomic DNA, complete sequence, section
1/4.
ACCESSION AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB038345 AB038346
AB038347 BA000008
AP002545.1 GI:8978373
VERSION
KEYWORDS
SOURCE Chlamydia pneumoniae (strain:J138) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS Shirai.M., Hirakawa.H., Kimoto.M., Tabuchi.M., Kishi.F., Ouchi.K.,
Shiba.T., Ishi.K., Hattori.M., Kunari.S. and Nakazawa.T.
TITLE Comparison of whole genome sequences of chlamydia pneumoniae J138
from Japan and CW0129 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
REFERENCE 20330349
2. (bases 1 to 300050)
TITLE Direct Submission
AUTHORS Shirai.M.
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
AB038345-AB038347: Submitted (14-Feb-2000)
AB036071-AB036078: Submitted (18-Dec-2000).
FEATURES
source
1. 300050
/organism="Chlamydia pneumoniae"
/strain="J138"
/db_xref="taxon:83558"
/note="synonym:Chlamydia pneumoniae"
```

```
gene
CDS
complement(1..282)
/gene="Cpf0001"
complement(1..282)
/gene="Cpf0001"
/codon_start=1
/transl_table=11
/product="C7001 hypothetical protein"
/protein_id="BAa98211.1"
/db_xref="GI:8978374"
/translacion="MLGKIRIGLSLIYILCALNVGLIGITHNKNIITAKLCGGVSTP
AIOIYIIGIAGVLCISLFCFCFSKRSRHSBGDCSSGGCSHSHSDKN"
573..878
/gene="gatc"
573..878
/gene="gatc"
/codon_start=1
/transl_table=11
/product="Glu-tRNA Gln amidotransferase (C subunit)"
/db_xref="GI:8978375"
/translacion="MEQPHLDREELILLAKASALQISELLIOEYOTSASVITSNKA
LAIEIDDADESCSELFMHVNVNEDREDSVTSDFNREELRNVPESLGLVYPAVIA"
895..2373
/gene="gata"
895..2373
/gene="gata"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/protein_id="BAa98213.1"
/db_xref="GI:8978376"
/translacion="MYRSALFLAAVTLGELTATGVQVHFFHRIEAEQVGATISL
KEQALEQBELIDKRSRGEPLGAGVPGIKDNIHVGTIKTTGASVLENYOPED
ATVVERIKKEDGIILGLNMEFAMGSTTLTSAFPHNTNPMDLTCAVSLEGGSAAV
SARCPVALGSDTGSIRQPAFCGVVFGKPSYGVASRYGLVAFASISDQIGPLANTY
EDVALMPDVEGSRDPKDATREFRDSFMSKLTSTVPKIVGAFRTFLEGLRDIREN
FSLAIFPEGEHLVDELIDILSHAVSIYIILASAEATNLARFEGVYGRSPQAT
ISOLDLSRGEGFEKVMRILIGVYISAEQONYVKKATVAPARKIYKARTAREKC
ELIAMPVCSAPFELGETLDPVTLGLDVIYVAMLAUPLAIAPSGFSKRGELPLGLQ
IIGQGDQDVQCVGVSFQEHNAQIKQLFSKRYAKSVVLGGOS"
2370..3836
/gene="gatb"
2370..3836
/gene="gatb"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (B subunit)"
/protein_id="BAa98214.1"
/db_xref="GI:8978377"
/translacion="NSAYYADWESYIGLEVHVELNTASKLESSALNREGDEPNRIST
VCTGLPSGLPYLNQSAVERKAVLFGAVGEISLSRFDRKSTFYPPDSRNFOITQEH
PIIRGRIRKAIYQGEERYELAQTHIEDDAQMLKHFGEAGVDVNRAGVPLEIVSKP
CMFCEPDAVAYATSLVSLDYIGISDCNMBEGSIRFDVNAVVRPSPGSELNKKYIKN
NMSFAFMAQALAEAKROIDEYLNQNPDKPLVIPAAYNVDRPEKKTVLNRKESAE
DYKAFPEPDLPTLQTESYIERIKRTELPDYKHAHYIOEGYSEDIASILISXNI
ATFEVACKDKNFRSLNMYSVERGSGCKLGYKLPSGSGIFPBGVADLVNADQGI
TKIARKEADLMMESPGKRNPEILKEKELLPSDBEQLQKILAEVVLANPESTIVDK
NGKTVALGVLQIMKRTGAKPARKRVNELLLELDKG"
4127..6895
/gene="pmp_1"
4127..6895
/gene="pmp_1"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family"
/protein_id="BAa98215.1"
/db_xref="GI:8978378"
/translacion="MRFSLGPEPLVFSFTLSLVDFTSLSATTISLTPEDSPHSONA
ESRYVQAGDYSLNGDYSLNVNDSANLAKNCFVNTSSYFACNHHHLVNNINSSGT
TRGAVLCCQDPQATARSSTLSFTIOSPEDIKQGLYKNAFLMLNNLVNREON
QSKTKGAISGANVIVGNVDSVFYQNAATFGGAIHSSGQLQIAVNOAEIRFQNTA
KNGSGALYSDGDIDIDONAVVLFRENEALTAIKGGAGVCCLPISGSSFPVIVTES
```

```

DNKQVFNHNSIMGGAIAIYARKLSISSGGPTLFINNISYANSONLIGAIADTGGEI
SLNAEKGTITFOGRTSLPFLNGHLLONAKFKILOARNGXYTEFPDITSEDSSTO
LNIQDPKNEKYCTGTLTFSCEKSLANDPRFKSTIPONVALSGITLVIRGAEVTVSK
FTOSPGSHVLDTGTLTKITLASKEDIALITGLAIDIDSSSSTAIVANIKANKOISVTD
SILISTGNAEYEDLRNRNSOTPLSLPEAGAGSVTVAGDFLPVSPHAGGMNKL
AMWGTGKRVGEFMDKINRKPREKGNLVPNINGNADVRLSMQVOETHASLQTD
RCGLMDIGNEFHVASSEDNIYRHNNGGVVLVNNNEITPKHYTSMFSQLFESRDY
AVSNNEYRMVLSYLYOYTTSLGNIFRYASRNPNVNGILSRELFONPLIMIPHELCAV
GHATNDKTDYANFPMYKNMNRNNCAIECGSGMPLIFENGSLFOGALIPMKQLQVY
AYGDFRETTADRRFNSGLTISVPLGIRFKKALSDQVLYDEFSFTIPDIFKDP
SCERALVTSGDSWLVPAHVSRHAFVSGTGRYHENDYTELLCRGSIETRPHARNINI
NCGSKTRF"
/gene="CPj0006"
/complement (7138. .7293)
/gene="CPj0006"
/complement (7138. .7293)
/gene="CPj0006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLSEWVLVGVSPETTRSTPEKANDLPKDSR
NRLESL"
/gene="CPj0009"
/6605..10499
/gene="CPj0007"
/7605..10499
/gene="CPj0007"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98217.1"
/db_xref="GI:8978380"
/translation="MRKRLIAIVLIALSIILAGSVLLTVAIPGLSSVISPAGM
AALGCVMLAGIDVLKRRKREYIVLASVTTGPMYSPSGSISGADSTISLPTYL
LDEGHQSMRKRLIALIVLIVLSIILAGVLLTVAISGLSVISPRARMACALGC
VMAIGIDVLKRRKREYIVLASVTTGPMYSPSGSISGADSTISLPTYL
QSMRKRLIALIVLIVLSIILAGSVLLTVAIPGLSSIISPAEWGACALGCVMLAG
IDVLLKREYIVPAPIPEEVVIDIDIESIRLOEAPALARLPEEMSAEFGYIKV
VESHENMSLPEVDHGHEKTKHOIRVRSLSKAMPPEFLIRLFEEMSAEFGYIKV
KRILDIATTLVERKILTEOLEERNNLKRAFSYODSIFKKIIDNEKILAMKMLISK
ICRFTIIFENHEHGVAKSLHKNAVILEKVIYRSLOKSYROI GMSAKMKILAGNPF
SLEDNKTIIMKEHAEMLESLSYRKVFILSDENVDPSPDKKMDLSIGIPRDLSE
ISRDQWOKKAHLKHESLYTOARDRLTQSSKENOKLEKAEQETISWEVKKFEI
EYQERIOALQKLPNILEBEETTGQETVPTVQCTAASDLTILGRIEVSSEDN
QWQESCVAKVRSHVEWSEWVKOEYGPKRKEQDOMGSLERFTEHIELEWLOQDS
KHLSTFKYVNNKREYQAKFRLKLVESLDEGLAOTESSELTOEELPIATRGALY
KAVFGSLCCALASKAKPYFEDPREDDSDTOLRALTILIOAKASLEDEIKRSENL
NDIAERRLIKESKOTFERAGLGVLRETAVESTYDLRSLTNWEGTSPSEKYPFMYL
NYNNEKRAKRIYEMTORRYDFFKALFAMQFNEDALLOEELSIQAFSE"
/10975..11668
/gene="CPj0008"
/10975..11668
/gene="CPj0008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98218.1"
/db_xref="GI:8978381"
/translation="MISGLFLVLRREYTVRSEIIPRGVSVTPSEEPALKEAKOEPE
TKIILDRPKELDLDYIOEYFACLERLKDPKRYLDELAKKALFVPPVKEKDM
SEFLIDIOVLNFEAYVHOCPLNINAYEIFSSELDDYICAGCGILPBGDNARAD
LKRSYKEYMDRMRYTKWSWEASVMLDHSYGARLFRKKAAGVLEESVYKILFYSYD
AFYECEKAKIORDGRFKWL"
/11815..13122
/gene="CPj0009"
/11815..13122
/gene="CPj0009"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98219.1"
/db_xref="GI:8978382"
/translation="WMIDRYADKFLIREKEKMERHELPHATWVRKASGHAYAKAKAA

```

```

PEKRSNENOKRVKVEWWSKGLAEFNQDSRRARERLRELQTLYPEVSEVYLER
QNTKRVNLENIYADIEKRYHHCVROEHYMKVEVNEKXEYENENEKVLASAEVSECQ
RLEMDLETFVSKRLTKEESVEMKFDATKGNVLSVTRRLILCEDABEMTFRIE
ELEMILRMVVELLFTMKNTPEKASLOYNCKEMAKYPOCKESPTVRSOERLELN
OOLQATVYNNCOERLOGFSDLSKVTICDHLREOMKHEVOGLNFINELLWGAELF
TOARLDAVATVPMYFLOYINIKREKRSQWMAKTERYRRIROAFQVYMKEDLAE
TLKEDEKEDVLRDMDLDEDERKNRRLITCKIAAQQRVAGF"
/gene="CPj0010"
/13434..15758
/gene="CPj0010"
/13434..15758
/gene="CPj0010"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98220.1"
/db_xref="GI:8978383"
/translation="MLLISGALFLFTLIGIPGLSAISFGGLGIGLSALGVLMTISGLC
LLVREIPTVPEEIPESVSLAPSEEPALQAQKTLAQPRELDQDLPDIOEVFACLR

```

alignment_scores:

Quality:	4746.00	Length:	928
Ratio:	5.114	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-428-122-2 x AP002545 ..

Align seg 1/1 to: AP002545 from: 1 to: 300050

```

1 MetlySerSerPheProlyPheValPheSerThPheAlaIlePhePr 17
|||||
21400 ATGAGAGCTTCCTTCCCAAGTTGATTTTCACATTGGCATTTTCC 21449
17 cLeuSerMetIleAlaThiGluThyValLeuAspSerAlaSerPhea 34
|||||
21450 TTTGTATGATGATTGTCACGACAGACTTTTGATTCAATCGCAGATTGC 21499
34 spcLyAsnLyAsnGlyAsnPhSerValArgLuserGlnGluAspAla 50
|||||
21500 ATGCAATATAAATGTAATTTTTCGTCGAGACGTAGCAAAATGCT 21549
2150 GlyThThThTytleuPheLySglAsnValThrLeuGluAsnIleProgl 67
|||||
21550 GGAACCTACCTACCTATTTAAGGAAATGCACCTAGAAAATATTCGCG 21599
67 yThrGlyThrAlaIleThrLySerCySephSnsnThrLySglAspL 84
|||||
21600 AACAGGACAGCAATCAAAAAGCTGTTTAACAACACTAAGGCGCAT 21649
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
|||||
21650 TAGACTTTCACAGGTAAAGGGAACCTCTATTGTTCCAAACGCTGATGCA 21699
101 GlyThValAlaGlyAlaAlaValAsnSerSerValValAspLySerTh 117
|||||
21700 GGGACGTGTGACAGGGGCTGCTGTTAAACAGCAGCGTGTAGATAAATCTAC 21749
117 thrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
|||||
21750 CACGTTTAAAGGGTTTCTCGCTATCTTTATTATGCGTCTCTGGAAGTT 21799
134 erIleThrThrGlyLySglAlaValAserCySerThGlySerLeuSer 150
|||||
21800 CGATACACTACCGGCAAGAGCGCTTACTCTCTACAGGATAGCTTGACT 21849
151 IeuthrLyAsnValSerLeuLeuPheSerLyAsnPhSerThrAspAs 167
|||||
21850 TTGAC.AAAAATGTCAGTTGCTTTCAGCAAAAACCTTTTCAACGAGATAA 21898
167 nGlyGlyAlaIleThrAlaLyThrLeuSerLeuThrGlyThrThrMetS 184
|||||
21899 TGGCGGTGCTATCACCGCAAAAACCTTTTCATTACAGAGGACTCAAAATGT 21948

```

184 erAlaLeuPheSerGluAsnThrSerSerLysLysGlyAlaIleGln 200
 |||||
 21949 CAGCTGCTTTCTGAAATATCTCTCAAGAAAGGCGGAGCATTCAG 21998
 201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSe 217
 |||||
 21999 ACTTCGAGTCCCTTACCATTAAGGAAACCAAGGAGAGTCTCTTTTC 22048
 217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluIaSerV 234
 |||||
 22049 TGACAATATCTCTCGATCTCGAGCTGCAATTTTACAGAACCTCGG 22098
 234 aThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
 |||||
 22099 TGACTATTCTTAATTAATGCTAAAGTTCTTTATGACAAATAGCTCA 22148
 251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCysAl 267
 |||||
 22149 GGACGCGAGCTCTCAACACGCGGAGATATGTCAGAGAGTCTATCTGCG 22198
 267 aTyrlLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
 |||||
 22199 TTAATAAACTAGTACAGATTAAGTCAACCTGCACTGGAATCAGATGT 22248
 284 euLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTyrlVal 300
 |||||
 22249 TACTCTTCAGACAAATACATGACAAACGCGGAGGAGCTATCTATGTG 22298
 301 LysLysLeuGlnLeuAlaSerGlyLysLeuThrLeuPheSerArgAsnSe 317
 |||||
 22299 AAAAAGCTCGAAGCTGCGTCCGAGAGACTTACCTATTCACATAAATAG 22348
 317 rValAsnGlyLysThrAlaProLysGlyAlaIleAlaIleGlnLys 334
 |||||
 22349 TGTCAAATGAGAGTACAGCTCTAAAGGTGAGCCATACCTATCCAGATYA 22398
 334 eArgLysLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
 |||||
 22399 GTGGGGAATTAAGTTATCCGCCATAGTGTGTCATTTCTTTTACGG 22448
 351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
 |||||
 22449 AATACATGACTCTACTACTCTCGGCGCAATAGAGTACTATGCACTT 22498
 367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIleArgAlaI 384
 |||||
 22499 AGGACCAAGTCCAAAGATGACAGCTTTCGCTGCTGCTGTAGAGCCA 22548
 384 leTyrlPheTyrlAspProIleThrThrGlySerSerThrThrValThrAsp 400
 |||||
 22549 TCTACTTCTAGATCCCATATACAGAGATCATCCACAACAGTTACAGNT 22598
 401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrlThrG 417
 |||||
 22599 GTCTTAAAGTTAATGAGACTCCGCGAGATCTGCACATCAATATACAGG 22648
 417 yAsnIleIlePheThrGlyLysLeuSerGluThrGluAlaIleAspS 434
 |||||
 22649 GAACATATCTTCACAGAGAAAGTTATCAGACACAGGCCCCAGATT 22698
 434 eTyrlAsnLeuThrSerLysLeuGlnProValThrLeuSerGlyLys 450
 |||||
 22699 CTAAATATCTTACTTCGAGAGTACTACAGCTGTAACCTTTCAGAGGT 22748
 451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
 |||||
 22749 ACTTACTTTTAAACATGAGTACTCTGACACTCAGGCAATTCACATCA 22798
 467 nGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlnProA 484
 |||||
 22799 ACAGGCAAGATTCCTGCTCGAATGAGAGCTGAGAACTATTAACCTG 22848
 484 laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500

|||||
 22849 CTGATTAACACCATTAACAAATTTGGTCAATTAACATCACTTATATGAC 22898
 501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
 |||||
 22899 GGTGCAGAGAGCCAAATAATAGAACCAAGACTTACGTCAAAAATCTGAC 22948
 517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrlGlu 534
 |||||
 22949 TTTATCTGGAACCATCACTTATATGAGCCGAGCGGAGCTTTATAGAAA 22998
 534 snHisSerLeuArgAsnProGlnSerTyrlAspIleLeuGlnLeuLysAla 550
 |||||
 22999 ATCATAGTTTAAGAAATCCAGAGCTTACGACATCTTAGAGCTCAAGCT 23048
 551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
 |||||
 23049 TCTGAGACTTAACAAACACCGAGTACTCCAGATCTTATATGGGTGA 23098
 567 uLysPheHisTyrlGlyTyrlGlnGlyThrTyrlGlyProIleValTyrlGly 584
 |||||
 23099 GAAATTCATTAAGCGCTATCAGGGAACCTTGGGCCCAATTTGTTGGGGA 23148
 584 hArgAlaSerThrThrAlaThrPheAsnThrThrLysThrGlyTyrlIle 600
 |||||
 23149 CAGGCGCTTCTACACTGCAACCTTCACTGAGCTTAACATGGCTATATT 23198
 601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuThrAsnAl 617
 |||||
 23199 CCTAATCCGAGCGGTATCGGCTCTTAGTCCCTAATAGCTTATAGGAATGC 23248
 617 aPheIleAspIleSerSerLeuHisTyrlLeuMetGluThrAlaAsnGluG 634
 |||||
 23249 ATTTATAGATTAATGCTCTCTCATTAATTAATGAGAGCTCAAGGAAG 23298
 634 LysLeuGlnGlyAspArgAlaPheTyrlCysAlaGlyLeuSerAsnPheP 650
 |||||
 23299 GGTTCGAGGAGACCGGCTTTTGGTGCTCGATTAATTAATCTTCTTC 23348
 651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGly 667
 |||||
 23349 CATTAAGATAGTACAAACACACGCGGCTTCCGCAATTTGAGTGCGCG 23398
 667 yTyrlValIleGlyLysLeuHisThrCysSerAspLysIleLeuSerA 684
 |||||
 23399 TTATGTCATAGGAGAACTTACATTAATTTGTCAGATTAATTTCTTAGTG 23448
 684 laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrlPheValAlaLys 700
 |||||
 23449 CTGATTTTGTCAAGCTCTTGAAGAGATGAGACATTAATTTGATAGTAAG 23498
 701 AsnGlnGlyThrValTyrlGlyGlyThrLeuTyrlTyrlGlnHisAsnGluTh 717
 |||||
 23499 AATCAAGGTACAGTACGAGAGAACCTCTTATTAACAGCACAAACAAAC 23548
 717 rTyrlIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrlValP 734
 |||||
 23549 CTATATCTCTTCTTTCGAACTACAGGCTTGTGTTGTTCTTAATGTTTC 23598
 734 roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrlThrHisThr 750
 |||||
 23599 CTACAGAGATTCCTGTTCTCTTTCAGAAACCTTAGCTACACCACTACG 23648
 751 AspAsnAspLeuLysThrLysTyrlThrThrTyrlProThrValLysGlySe 767
 |||||
 23649 GATTAAGATGTGAACCAAGTATACAACTATTCCTACTGTTTAAAGGAAG 23698
 767 rTyrlGlyAsnAspSerPheAlaLeuGlnPheGlyGlyArgAlaProIleG 784
 |||||
 23699 CTGGGGAATAGTATGTTGCTTATGAAATTCGTTGGAAGAGCTCGCATTT 23748
 784 yLeuAspGlnSerAlaLeuPheGluGlnTyrlMetProPheMetLysLeu 800
 |||||

23749 GCTTAGATGAAGTCTCTATTGACAGACATGCGCTCATGAATTG 23798
 801 GlnPheValIYrAlaHisGlnGluGlyPheLysGlnGlyThrGluAl 817
 23799 CAGTTTGCTATGCACATCAGGAAGTTTAAAGAACGGGGAACGAAGC 23848
 817 AARGGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
 23849 TCGTAATTTGGAAGTAGCCGCTCTGTAATCTTCCCTTACTTACGGGA 23898
 834 leArgPheAspLysGlnSerAspGlyAsnAspAlaThrTyrAsnLeuThr 850
 23899 TCCGATTTGATAGGAATCAGACCTGCCAAGATGCACACTACAACTTAAT 23948
 851 LeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
 23949 CTGGTGTACTGTGATCTTGTCTGTGTAACCCCGACTTACGACACAC 23998
 867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884
 23999 ACTGCAATTAGCGGTGATCTTGGAAACCTTCGGTACGAATTTGGCAA 24048
 884 rGGAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
 24049 GACAGCTTTGCTCTTCCTGTCAGGGAACCAATTTTCTTAACTCAAT 24098
 901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
 24099 TTTGAGCGCTTTAGCCAAATTTCTTTGAAATGCGTGGGTATCTCCGA 24148
 917 nTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
 24149 TTACAATGTAGACTTAGAGCAAAATACCAATTG 24182
 seq_name: gb_ba1:AB033795
 seq_documentation block: 2241 bp DNA BCT 31-OCT-1999
 LOCUS AB033795 Chlamydomophila pneumoniae pmp-3.2 gene for polymorphic outer
 DEFINITION membrane protein G family, complete cds.
 ACCESSION AB033795
 VERSION AB033795.1 GI:6172316
 KEYWORDS POLYMORPHIC outer membrane protein G family.
 SOURCE Chlamydomophila pneumoniae (strain:J138) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE 1 (sites)
 AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
 Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
 Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,
 Ishii,K., Shiba,T., Hattori,M. and Kuhara,S.
 TITLE Comparison of outer membrane protein genes omp and pmp in the whole
 JOURNAL genome sequences of Chlamydia pneumoniae isolates from Japan and US
 REFERENCE Unpublished (1999)
 AUTHORS 2 (bases 1 to 2241)
 TITLE Hirakawa,H. and Shirai,M.
 JOURNAL Direct Submission
 REFERENCE Submitted (23-OCT-1999) to the DDBJ/EMBL/GenBank databases. Hideki
 AUTHORS Hirakawa, Kyushu University, Graduate School of Genetic Resources
 TITLE Technology: 6-10-1 Hakozaeki, Higashi-ku, Fukuoka, 812-8581,
 JOURNAL Japan (E-mail:hirakawa@grt.kyushu-u.ac.jp,
 URL:http://www.grt.kyushu-u.ac.jp, Tel:81-92-642-3043,
 FEATURES
 source
 1. .2241
 Location/Qualifiers
 /organism="Chlamydomophila pneumoniae"
 /strain="J138"
 /db_xref="taxon:83558"
 /note="Synonym:Chlamydia pneumoniae"
 1. .2241
 /gene="pmp-3.2"
 1. .2241
 /gene="pmp-3.2"
 CDS
 /gene="pmp-3.2"

/note="containing frame shift"
 /codon_start=1
 /transl_table=11
 /product="polymorphic outer membrane protein G family"
 /protein_id="BA85948.1"
 /db_xref="gi:6172317"
 /translation="MSALFSENTSSKKGAQIOTSDAITITGNGEVSFSDTSSDGA
 ALFTBASVTISNNKVSFTDNKVTGASSSTGTGDSGACIAKYSTDTKTYLGNOML
 LEFNNSTTACGAIYVKRLASGGLLFSRNSVNGGTAIGALIAIEDGELSAD
 SGDIYFLGNVTSTPGINRSSIDLGTSAKMTALRSAGRAIYFDPITTGSSSTTV
 VLKVEITPADSALOYTGNIIFTEGKLESETEAADSKNLTKILOPVTLSGGTLSKHG
 TLQOAFQOADSRLQEMDVTLEPADSTRTNNVINSIDAKAKIEFKASKNT
 TLSGITLIDPTGTFEYENHSIRNPOSYDILEKRSQVTSYTAIYDPIDMEKTHYGTO
 GTWGPVWGTASTTATFNTKIGTIFNPEITGSLVPSNLSMNAITDISLHYMETAN
 EGLQDRAFWCAGLSNFEHDKSTRTGRGFRHLSGVYIGNLTCSDKILSAFCOLF
 GRDRDYFAKNQGVYGGITLYQHNETYISLPCRLRPSLSYETETPLVFSGLSYT
 HTDNDLKRTYTPYTVKSGMNDSEFALFEFGRAPICLDESALFEQVYPMKLOFVVAH
 QEGFKEQGTFAREFGSSRLVLAIPIGIRDKESDQDQATYNTLTGYTMDLVSNDC
 TTTLRISGDSKRTGTMALRALVLRAGNFCFNSNFARSFSEFLRGSSRNINDL
 GAKYQF"
 BASE COUNT 660 a 474 c 478 g 629 t
 ORIGIN
 alignment_scores:
 Quality: 3865.00 Length: 746
 Ratio: 5.181 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-428-122-2 x AB033795 ..
 Align seg 1/1 to: AB033795 from: 1 to: 2241
 183 MetSerAlaLeuPheSerGluAsnThrSerLysGlyAlaAl 199
 1 TATGCACTCTGTTCTGAAAATACCTCTCAAGAAAGCGGAGCAT 50
 199 eGlnThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerP 216
 1 TCGAGCTTCGATGCCCTTACCATCTGGAACCAAGGGGAAGCTCTT 100
 51 TCGAGCTTCGATGCCCTTACCATCTGGAACCAAGGGGAAGCTCTT 100
 216 hSerAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAl 232
 101 TTTCTGACATACCTCTTCGATCTCGAGCTCGCAATTTTACACAACC 150
 233 SerValThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLys 249
 151 TCGGTGACTATTCTTAATATGCTAAAGTTTCCTTATTGCAATTAAGT 200
 249 lThrGlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleC 266
 201 CACAGAGGAGGAGCTCTCAACACGGGGGATATGCAAGAGGTCTATCT 250
 266 yAlaIleTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGln 282
 251 GTCCTTATAAACTAGTACAGATACTAAGTCAACCTCAGCAATCAG 300
 283 MetLeuLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTy 299
 301 ATGTACTCTTCAGCAACAATACATCGACAAGGGGAGGAGGTATCTA 350
 299 rValLysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArg 316
 351 TGTGAAAAGCTCGAACCTGCTCCGGAGGACTTACCTATTCAGTGA 400
 316 snSerValAsnGlyGlyThrAlaPolysGlyGlyAlaIleAlaIleGlu 332
 401 ATAGTGCAATGAGAGTACAGCTCTTAAGGTGAGGCATAGCTATGAA 450
 333 AspSerGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPhe 349
 451 GATAGTGGGGAATTGAGTTTATCCGCCGATAGTGTGACATTGTCTTTT 500

349 uGlyAsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleA 366
|||||
501 AGGGAATAACACTCTACTACTCTCCGGACGAAATGAAAGTAGTATG 550
366 sPLeuGlyThrSerAlaIleuMetThrAlaIleuArgSerAlaIleuGlyArg 382
|||||
551 ACTTAGGACACAGTGCNAAGATGACAGCTTGGCTTCTGCTGTGTA 600
383 AlaIleuThrPheTyraSPProIleThrThrGlySerSerThrThrValTh 399
|||||
601 GCCATCTACTCTATGATCCATACACAGATCATCCACACAGGTAC 650
399 rAspValIleuLysValAsnGlyThrProAlaAspSerAlaIleuGlnTyT 416
|||||
651 AGAGTCTTAAAGTTAATGAGACTCCGGCAGATTCTGCCTACAAATTA 700
416 hrGlyAsnIleIlePheThrGlyGlyLysLeuSerGlyThrGluAlaIle 432
|||||
701 CAGGGAACATATCTTCACAGAGAAAAGTTATCAGACAGAGGCCCA 750
433 AspSerLysAsnLeuThrSerLysLeuGlnProValThrLeuSerG 449
GATCTTAAATCTTACTTCGAGCTACTACAGCCTGTAACCTTTTCAG 800
449 yGlyThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheT 466
|||||
801 AGGTACTTATCTTTAAACATGAGTACTCTCAGACTCAGCATTCA 850
466 hrGlnGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGly 482
CTCAACAGGCGAGATTCTGCTCGAAATGAGACGTAGAACTACTCA 900
483 ProAlaAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSer 499
|||||
901 CCTGCTATCTACACCATAAACATTTGGTCACTTAACATCACTTCAT 950
499 eAspGlyAlaLysLysAlaLysIleGlyThrLysAlaThrSerLysAsn 516
|||||
951 AGACGTGCAGAAAGCAAAATGAAACCAACCTACGTCNAATC 1000
516 eThrLeuSerGlyThrIleThrLeuAspProThrGlyThrPheTyT 532
|||||
1001 TGACTTTATCTGGAACCATCTTATTTGACCCGACGCGCTTTAT 1050
533 GluAsnHisSerLeuArgAsnProGlnSerTyraSPIleuGlnLeuTy 549
|||||
1051 GAAATCATATAGTTAAGAAATCTCAGTCTACGACATCTTAGAGCTCA 1100
549 sAlaSerGlyThrValThrSerThrAlaValThrProAspProIleMet 566
|||||
1101 AGCTTCTGGAATCTACACAGCAGCAGTACCTCAATCTTAATAG 1150
566 LysGlyLysPheHisTyGlyTyGlnGlyThrTPGlyProIleValTyT 582
|||||
1151 GTGGAATTCATTCATACGGCTATCAGGGAACCTGGGCCCAATGTTG 1200
583 GlyThrGlyAlaSerThrThrAlaThrPheAsnThrPheLysThrGlyTy 599
|||||
1201 GGGACAGGGGCTTCTACGACTGCAACCTTCACAGGAGCTAAACGGCTA 1250
599 rIleProAsnProGluArgIleGlySerLeuValProAsnSerLeuTyP 616
|||||
1251 TATCTTAATCCGAGCGTATCGGCTTTAGTCCCTAATAGCTTAAGGA 1300
616 snAlaPheIleAspIleSerSerLeuHisTyTLeuMetGlyThrAlaAsn 632
|||||
1301 ATGCATTTAATAGATTAGCTCTCTCATTAATGATGAGACAGCAAC 1350
633 GluGlyLeuGlnGlyAspArgAlaPheTyPcysAlaGlyLeuSerAsnPh 649
|||||
1351 GAAAGGTTGCAGGAGACGCTTTTGGTGTGCTGATTAATTAATCT 1400

649 ePheHisLysAspSerThrLysThrArgGlyPheArgHisLeuSerG 666
|||||
1401 CTTCATTAAGATAGTACAAAACACAGCAGCGGCTTCCGCAATTTGAGTG 1450
666 LysGlyTyValIleGlyLysLeuHisThrCysSerAspLysIleLeu 682
|||||
1451 GCGGTTATGTCATAGAGAAACCTACATCTTGTAGATTAACATTTCT 1500
683 SerAlaIlePheCysGlnLeuPheGlyArgAspArgAspTyPheVal 699
|||||
1501 AGTCTCATTTTGTACGCTCTTGGAAAGATGAGACTACTTTGTAGC 1550
699 aLysAsnGlnGlyThrValTyGlyGlyThrLeuTyTyGlnHisAsn 716
|||||
1551 TAAGAATCAAGTACAGCTACGAGGAACCTCTATTAACAGCAACG 1600
716 LuThrTyIleSerLeuProCysLysLeuArgProCysSerLeuSerTyT 732
|||||
1601 AAACCTATCTCTCTCTCTTGCAAACTACGGCCTTTGTGTCTTAT 1650
733 ValProThrGluIleProValLeuPheSerGlyAsnLeuSerTyThr 749
|||||
1651 GTTCTCAGAGAGATCTCTGTTCTCTTTTCAGAAACCTTAGCTACACCA 1700
749 sThrAspAsnAspLeuLysThrLysTyThrThrTyProThrValLys 766
|||||
1701 TACGATACAGATCTGAAACCAAGTATACACATATCTACTGTTAAG 1750
766 LysETrpGlyAsnAspSerPheAlaLeuGlnPheGlyGlyArgAlaPro 782
|||||
1751 GAACCTGGGGAATGATAGATTGGCTTTAGAAATTCGTTGGAAGGCTCCG 1800
783 IleCysLeuAspGlnSerAlaLeuPheGlnGlnTyMetProPheMet 799
|||||
1801 ATTTGCTTAATGAAAGTCTATTTGAGCAGTACATGCCCTTCATGAA 1850
799 sLeuGlnPheValTyAlaHisGlnGlyPheLysGlyGlnGlyThrG 816
|||||
1851 ATTGCACTTTGTCTATGACATCAGAAAGTTTAAAGAACAGGAAACG 1900
816 LuAlaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIle 832
|||||
1901 AAGCTCTGAATTTGGAGTAGCCGTTCTGTGAATCTTGCTTAACCTATC 1950
833 GlyIleArgPheAspLysGlnSerAspCysGlnAspAlaThrTyraSP 849
|||||
1951 GGGATCCGATTTGATTAAGAAATGAGCTGCCAAGATGCCACGTACATCT 2000
849 uThrLeuGlyTyThrValAspLeuValArgSerAsnProAspCysThr 866
|||||
2001 AACCTGTTATATCTGTGATCTGTGTTAGTAACCCGACGTACGA 2050
866 hrThrLeuArgIleSerGlyAspSerTyPlyThrPheGlyThrAsnLeu 882
|||||
2051 CAACACTGCCAATTTAGCGGTATCTTGGAACCTTGGTACGAATTTG 2100
883 AlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsn 899
|||||
2101 GCAGACACACTTTAGTCTTCTGTCGAGGGAACCATTTTGTCTTAATCT 2150
899 rAsnPheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSer 916
|||||
2151 AAATTTGAGCCCTTTAGCCAATTTCTTTGAAATTCGTGGCGCACTC 2200
916 rGAsnTyraSPValAspLeuGlyAlaLysTyGlnPhe 928
|||||
2201 GCATTTACAATGTAGACTTAGAGCAAAATACCAATTC 2238
seq_name: gp_pat: A81829
seq_documentation_block: 2815 bp DNA
LOCUS A81829
DEFINITION Sequence 3 from Patent WO9858953.

21-JAN-2000


```

1400 CCTGTTATATGATGCGGCACACAGCTTAAGAACAAGTACAGAGAG 1449
487 SerThrIleAsnLeuValIleAsnIleSerIleAspGlyAla 503
1450 GTCACCTTAACAGGCTCTTCCTTCTGACCTTTTAGCGGAGGTA 1499
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerg 520
1500 GAAAGTTGTAATGCTGCTTCGACAGCAAGTAATAATGAGCCCTTAG 1549
520 LysThrIleLeuLeuAspProThrGlyThrPheTyrgLysAsnIleSer 536
1550 GTCCGATCTCTTTGGATACCAAGGAGACTTATGAATATACAGAC 1599
537 LeuArgAsnProGlnSerTyrgAspIleLeuGluLeuLysAlaSerGly 553
1600 TTAGGAAAACACAGACTTTTCATTTGCGAGCTCTGCTCTGGGTAC 1649
553 rValThrSerThrAlaValThrProAspProIleMetGlyIulysPhe 570
1650 TGCAACACACTACAGATGTTCCAGCGGTTCTTACAGTACACACTCT 1699
570 sTyrgGlyTyrgGlnGlyThrTrpGlyProIleValTrp 582
1700 ACTATGGGTATCAAGGACTTGGGA...ATGACTTGGGTGATGATACC 1746
583 ..GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThr 598
1747 GCAAGCACCTCAAGACTAGACAGCGACTTACCTGACCAATACAG 1796
598 yThrIleProAsnProGluArgIleGlySerLeuValProAsnSerLeu 615
1797 CTACCTTCGAACTCTGAGCGTCAAGACCTTACTTCTTAATACCTTT 1846
615 rPasnAlaPheIleAspIleSerSerLeuHisTyrgLeuMetGluThr 631
1847 GGGGACTTTTTCAGACATCCAGGATTCAGAGTCTCATAGAGAGAGT 1896
632 AsnGluGlyLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSer 648
1897 GCTTGAGCTTTTCTTCAGATCGAGCTTCGCGCTCGCGAGTCCCA 1946
648 nPhePheHisLysAspSerThrLysThrArgArgLysPheArgHisLeu 665
1947 TTTCTTAATATAAGTAAAGAAAGGGAAGCAATACCGTCAATAAT 1996
665 erGlyGlyTyrgValIleGlyLysAsnLeuHisThrCysSerAspLys 681
1997 CTGGTGATATGCTATCGAGGTGCGCAAACTTGTGAAACCTTA 2046
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTrpPhe 698
2047 ATTACCTTTCCTTTTCCCACTCTTGTGACGATAAAGATTCTTACT 2096
698 lAlaLysAsnGlnGlyThrValTyrgLysThrLeuTyrgLysHis 715
2097 CGCTAAATATCATACTGATACCTATGACGAGCCCTCATATACCAAC 2145
715 snGluThrTyrgIleSerLeuProCysLysLeuArgProCysSerLeu 731
2146 .....ATTACAGAAATGATAGGTTCACTAGTCTCTTCTTATAT 2184
732 TyrValPro.....ThrGluIleProValLeuPheSerGlyAsn 745
2185 AAACCTCTGCTCTTGGAGTATAAACCCCTGTTTAGAAGGCGACCT 2234
745 userTyrgThrHisThrAspAsnAspLeuLysThrTyrgTrpThrTyrg 762
2235 CGCTATATAGCCACGTCAATGATCTCAAGCAAGATATACTGCTATC 2284
762 rOrthValLysGlySerTrpGlyAsnAspSerPheAlaLeuGluPhe 778

```

```

2285 CTGAGGTGAAGGTTCTTGGGGAATAATGCTTTAATCATGATGTTGGA 2334
779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTyrg 795
2335 GCTCTCTCTCTCTTCTTATCCGAATACCTGCATGTGTTTATACTATGC 2384
795 tProPheMetLysLeuGlnPheValTyrgAlaHisGlnGluGlyPhe 812
2385 TCCATATCATCAACTGAATCTGACCTATATACGTACAGAGAGCTTCGG 2434
812 LysGlnGlyThrGluAlaArgGluPheGlySerSerArgLeuValAsn 828
2435 AGAAGGTACAGAAAGATCTTTGATGACAGCAACCTCTTCAATTA 2484
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAsp 845
2485 TCTTTCCTATAGGGGTGAAGTGTGAGAACTTCTCATTTATGACTT 2534
845 aThrTyrgAsnLeuThrLeuGlyTyrgThrValAspLeuValArgSer 862
2535 TTCTTATGATCTGACTTATCTTATGTTCTCGATCTTATCCGAATGATC 2584
862 rAspCysThrThrThrLeuArgIleSerGlyAspSerTrpLysThrPhe 878
2585 CCAATGACATACAGCACTTGTATATCAGCGAGCCTCTGGGAACCTAT 2634
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHis 895
2635 GCCAATACCTTACGACAGCAGCGCTTCAGTCCGTCAGCAGCACTGAC 2684
895 eCysPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGlu 912
2685 CGCCTTCTCTCTTGTGAGTGTGCGGCACTGTGCTTGAAGTTC 2734
912 rGlySerSerArgAsnTyrgAsnValAspLeuGlyAlaLysTyrgLys 928
2735 GTGATCTCTACGAGATTATATGATGATCTTGGGGTAAGTTCATTC 2784

seq_name: gb_bai:AE002192

seq_documentation_block:
LOCUS AE002192 12676 bp DNA BCT 30-MAY-2000
DEFINITION Chlamydomophila pneumoniae AR39, section 25 of 94 of the complete
genome.
ACCESSION AE002192 AE002161
VERSION AE002192.2 GI:8163405
KEYWORDS
SOURCE
ORGANISM
Chlamydomophila pneumoniae AR39.
Chlamydomophila pneumoniae AR39.
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE
1 (bases 1 to 12676)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE
2 (bases 1 to 12676)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
JOURNAL
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
COMMENT
On Jun 1, 2000 this sequence version replaced gi:7189226.
FEATURES
Location/Qualifiers
1..12676
/organism="Chlamydomophila pneumoniae AR39"

```

```
gene /strain="Ar39"
CDS /db_xref="taxon:115711"
      /note="synonym: Chlamydia pneumoniae Ar39"
      complement(90..185)
      /gene="CP0300"
      complement(90..185)
      /gene="CP0300"
      /note="hypothetical protein; identified by Glimmer2;
      putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAF38157.1"
      /db_xref="GI:7189227"
      /translation="MFTSLGRSAIFYFTGSLKHLRGLGEFLFIQK"
      complement(275..1819)
      /gene="CP0301"
      complement(275..1819)
      /note="similar to GP:4376735; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="polymorphic membrane protein G family"
      /protein_id="AAF38158.1"
      /db_xref="GI:7189228"
      /translation="MTILRNFLTCSALFLALPAAQVYVYLHSDGNGAINKSLEPK
      ITCEPGTGYFLIDVIRISNKHQEDAGVFIRSGNLFPMGRGCFTHNLMTEGFG
      AINSNRYGDTTLTISNFSYLAFTSAPILPDGGGAIYSLGSVMIEENSEFTFCGNSSW
      SGAATYPTLLSKASRSRVSNSGRNLYFRNRYSGYGAISTHNLTITTPGSCFE
      NHRAYHDVNSNGGAIAPAGGSIISVNSGDLFGNTASDQGNTHNSIHLQSAQF
      KNRUVESGVYFDPIISHESHKITDLYINAPEKEYTEGTISFGCLDHEVCAL
      NLTSTILOVPTLAGGTLSDGVTLLQHLHSGFQKQASSTLTMSPGTLTLCGDAVYLNH
      LIIEPDNVPVPIRAEDKDALVLEKLVAFEAWSYVDPEQKEAFIPLLELQV
      SEDSLLETLERTQVTENDAVAGFMSLSEVEYPPSLDKDRRTITPKKYFLIWNP
      EITSTP"
      complement(2077..4926)
      /gene="CP0302"
      complement(2077..4926)
      /note="similar to GP:4376733; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="polymorphic membrane protein G family"
      /protein_id="AAF38159.1"
      /db_xref="GI:7189229"
      /translation="MIVLFCFYIDANSSILKNSITMKTSIPVAVSVYLAESCHQSL
      ANEELSPDPSFNGNIDSGTTPKTSATITYSLTGDFYFEPEKGTPLSDSCFKQITDN
      LTFLENGSLITGFLIDAGHAGAASTANKNLTFSGFSLSDSSPSTVYTGQGTL
      SSAGVNLNENIKRLVAGNFSTADGGAIRKAGFLLTGSGDLFENNSSSTKGAIAIT
      TAGARIANNQTVGRFLSNIASTSGAIDEGTISNNKFLYFEGNAAKTTGACATC
      KASGPELIISNNKTLIFASNVAETSGGAIRAKKIALSSGGFTEFLRANNAATPCKG
      AITSDAGSELISAEETGNITPYRNLTITGSDTPKRNAININGSKXETELRAANHT
      IFYFPTSEGTSSDYVLRKINNSAGALNPQGTIIFSEPTLTADLKAADNKSFTQ
      PVLSGGKLLDQGVLTLESTFSQAGSLGMDSTTLTSTAGSITITNLGINDSLG
      LKQVSLTPAKGASNRYVSGKLNLIIDIGNIEMSESHDQFLSLKITVADVDVNY
      DISSLIPVPAEDPNSEYFGQGMNNTDIDATNTKEATATWTGTFVPSERKCALY
      CNTLGVFTDIRLSIQOLVEIGATGMEHKGQFNSMTNLIHKTGDNKRKGRPRSGGY
      VTGSAHTPKDILFTFARCHLFARDKCFIANNNSRTYTGTLFFKHSHTLOPONTLR
      GRAKSESAIEKFPREIPLADVOYSESHSDKRMKTHYSLPESEGSNSNEC IAGIG
      LDLPVLSNPHPLEFTFI PQMKVENVVYVSQNSPESSSDGRGFSIGRLINLSIPGAK
      FVOGDIGSYDLDISGFVSDVYRNPNPOSTATLVMSPSMKIRGGLRSOAFLLRGSN
      NYVYNSNCELFGHYAMELRGSSRNRYNDVGTCLRP"
      complement(5026..7812)
      /gene="CP0303"
      complement(5026..7812)
      /note="similar to GP:4376729; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="polymorphic membrane protein G family"
```

```
gene /protein_id="AAF38160.1"
CDS /db_xref="GI:7189230"
      /translation="MKSQFWLVYSTLTACEFSCSTYPAATRNIGSPDSPTNRC
      TTPKNTTGGIDYLTGDTITQNLGDSALIKCFSPDTBESLSTRACKYSLSFLNIK
      SAEGALSVTTDKNLSTLTFGLAPSSVITTPSGKAVKCGGDTFDPNNGTILF
      KODYCEENNGAIISTRNLSLNKSTGISPEGRKSSATYATCGATGVDVNTNAPT
      LFSNNIAEAGAGAIINSTNCTITGNTSLVFSENSSTYATAGNGALSGADVTISGNOS
      VTFSSQAVANGAGIYAKKLLPLASGGGGISFSSNNIYOGTATAGNGALSIILAGCSL
      SAEAGDITFNGNAIYATTPORTKRSIDIGSTAKITNLRATISGHSIFPYDITANTTA
      DSTDTLNKADAGNSTDYGSSIVPSSGKISEDAKAVDNTSLTKOPVTLTAQNVYL
      KRGTITDKGFTQVRGSSVIMDAGITTLKASTEEYTLVGLSLPVSJLGGKRVYLAASA
      ASKNVALSGPILLDQGNAYENHDGKTQDFSVQUSALSTATTDDPAPVATPATPT
      HYGQGTGSMTWVDOTASTPKYKATLAWNTGYLPBERGQPLVPSLWMSFSDIQA
      IOGVYERALTLCSDPRGFWAGVANEFLDKDKGKRYRKHSGGAIIGAGQCTSEN
      ISFAFCQFGSDKDFLVAKNHTDYPAGAFYIOHTEGSGFICGLDLPKGSWSKPY
      LEGOLAYHVSNDLTKYKATYAYPEVGSWGNAPFMNMGASGSHPEYLHCEDTPAYPI
      KLNTYIRQDSFSEGTGGRSFDDSNLENLSLPIGVKREKSDCNDPSYDITLSVPD
      LTRNDPKCTIALVLSGASWETVANNLAQALQVRAGSHAFSPMFEVLGQPFVEVRS
      SRIYNVDLGGKFOF"
      complement(8170..8292)
      /gene="CP0304"
      complement(8170..8292)
      /note="hypothetical protein; identified by Glimmer2;
      putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAF38161.1"
      /db_xref="GI:7189231"
      /translation="WDLSSRWLNKLRKQOLDQNIKYDCKSYCLRGISQLIGWV"
      complement(8329..9444)
      /gene="CP0305"
      complement(8329..9444)
      /note="conserved hypothetical protein; identified by
      Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="AAF38162.1"
      /db_xref="GI:7189232"
      /translation="MEMWSPQPEQCHFDVYGSFLRPESITRARSDFEGRIYQEM
      RVEDDAIRNLIRKQTEAGLIFFTDGFERRYSWDPDMWGHGVDRSRDSNDPEIGY
      LKDKISVSKHPIEFHFEYKTFEKGNAKAKOTIPSPQFEMIFAFNLKTRKFTYPT
      NOELIDIDIVEFYROVIODLVYACGRNLOLDCACRLDIPRAPSVDSDRLOEEL
      EQLFMINHWKMDRPEDLPVSLHYCRDQYAEFRSRAYSIEPLRAKDDVDSYHXY
      WALDKISGSGKEPRLAYISGEKHVLCGLISSHSCIEDRDVAVSRITYEAASTIPLERLS
      LSPQGFASCSGDBRMTEDBQWKIAFVKEIAEWG"
      complement(9785..12571)
      /gene="CP0306"
      complement(9785..12571)
      /note="similar to GP:4376731; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="polymorphic membrane protein G family"
      /protein_id="AAF38163.1"
      /db_xref="GI:7189233"
      /translation="MKSSLHWFILISSLLALPLSINSAFAAWEINILGPNSESQPT
      YTPAQGTNMDGTYLNTGCVSITNAGSPITALASCEKETGNLSFGHGVOFLQNI
      DAGNACFTNTPAAKRLDSFGSEYSLSIQTTNATGAIKSTAGCSIOSSQYFQO
      NESNDNGALOGGSSISLISNNTLFAKNKATOKGALYSGYGTINNTLINSASPE
      AANNCGAIYTEASSFISNNKAISTINNSVATSTGAIYCSSISAKPVLUTSDNE
      LNFIGNRAIISGAIYTDNLVLSGGFTLRKNSAIDTAAPLGGALIAJDSGLSLISA
      LGDITFEGNTVYVKGASSQITRNSININTNNAKIVQLASQGNITYFYDPIITTSIT
      AALS DALNNGPDLAGNPAYOGTIVFSGERLSEAAEADNLKSTIOOPLLAGGQSL
      LKSGVTILVAKSEFSGSPSTILMDAGTTLTADGITLINLVANDSLKETKATKAKO
      ASQVITLGSISLSDPSGNVYEDVSNPNPOVSCITLTADDPANIHITLTDLAADPLEN
      PIHWYOGNNALSNQEDTATKSKAATITWTKGTGNPBERGRLIVANTLNGSYDVYS
      IQOLVARKROSQETRGITWCEGISNPFHKDSTYKINKFRIISAGYVAGATTTLASDL
      ITAAFCOLFGRDHFITKNKASAYASLHLQHLATLDRSLSLRLYLGSESESEDPVLD
```

AOISYISKNTMTYTTQAPKSSMYNDCALFLASLPHLALSHGFLHAFYPPK
 VEASVYHODSKERNITLVRSFDSGLINVSVPITGTFERSNERASRYEATVYVAD
 VYRKNPDCITALLINNTSMKTTGTNLSRQAGIRAGIFAFSPNLEVTNLSMEINGS
 SRSYNDIGKPF"

BASE COUNT 3580 a 2439 c 2890 g 3767 t
 ORIGIN

alignment_scores: Length: 950
 Quality: 2058.00 Gaps: 20
 Ratio: 2.831
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:
 US-09-428-122-2 x AE02192 ..

Align seg 1/1 to: AE002192 from: 1 to: 12676

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePhe 16
 |||||.....
 5026 ATGAATATGCAATTTTCCTGGTTAGTCTCTCTCGACATTGGCATGTTT 5075
 16 e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSer 30
 |.....
 5076 TACTAGTTGTTCCACTGTTTTCGCTGCAACTGCTGAAATATAGCCCTT 5125
 30 eAlaSerPheAspLysAlaValAsn...GlyAsnPheserValArgLys 45
 |||||.....
 5126 CTGATAGCTTTGACGAGTACTACACAGCAGCCTATACCTCTTAAT 5175
 46 SerGlnGluAspAlaGlyThrThrThrLeuPheLysGlyAsnValThrLe 62
 :|||:.....
 5176 ACG.....ACACTGGAATAGACTATACTCTGACAGGAGATATACCT 5219
 62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
 :|||:.....
 5220 GCAAAACCTT...GGGGAATTCGGCAGCTTTAACGAGGTTGTCTG 5266
 79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
 :|||:.....
 5267 ACACACGGAATCTTAAAGCTTTCGGGTAGGGGTACTCCTCTCTTCT 5316
 96 GlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
 :|||:.....
 5317 TTAATATATAGTCT...AGTCTGAGGCGCAGCAGCTT...TCGTATAC 5360
 112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIle 129
 :|||:.....
 5361 AACTGATATAAATCTGCTGCTAACGAGATTTTCGAGCTTACTTCTTAG 5410
 129 lAspProGlySerSerIleThrThr.....GlyLysGlyValAlaSer 143
 :|||:.....
 5411 CGGCCCTTCATCGTATACACACCCCTCAGGAAAGGTGCAGATTGAA 5460
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
 :|||:.....
 5461 TGTGGA...GGGGAATCTTACATTTGATTAACAATGCAATATTTATTAA 5507
 160 rLysAsnPheserThrAspAsnGlyAlaIleThrAlaLysThrLeu 177
 :|||:.....
 5508 ACAAGATTACTGTGAGAAATGGCGGAGCATTCTTACCAAGATCTTT 5557
 177 eLeuThrGlyThrThrMetSerAlaLeuPheserGlyAsnThrSerSer 193
 :|||:.....
 5558 CTTTGAACACGACGAGGATGATTTCTTTGAGGAGGATTAATCGAGC 5607
 194LysLysGlyAlaIleGlnThrSerAspAlaLeuThr 207
 :|||:.....
 5608 GCAACAGGAGAAAGAGTGGGCTATTGTGCTACTGCTAGCTGATGAT 5657
 207 eThrGlyAsnGlnGlyLysValSerPheSerAspAsnThrSerSer 224
 :|||:.....
 5658 TACAATATATACGCTCTCTACCTCTTCTCGAACATATATGCTGAGCTG 5707

224 eArgLysAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
 :|||:.....
 5708 CAGGAGAGCTATATAATACACAGGAACTGACATATACAGGAAATPAC 5757
 241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThrTh 257
 :|||:.....
 5758 TCTCTGATATTCTGAAATAGTGAGCA.....GGACCGC 5795
 257 rGlyAspMetSerGlyGlyAlaIleCysAlaThrLysThrSerThrApr 274
 :|||:.....
 5796 AGGA.....AATGAGAGAGCTCTT.....TCTGAGAGATG 5824
 274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
 :|||:.....
 5825 CCGATGTACCATATCTGGGAATCAGATGTAACTTCTCAGAAACCA 5874
 291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluAlaLeu 307
 :|||:.....
 5875 GCTGAGCTAATGGCGAGCCATTATATGCTAAGAGCTTACACTGCTTC 5924
 307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLys 322
 :|||:.....
 5925 CGGGGGGGGGGGGTATCTCTTTCTAACAATATAGTCCAGAGTACA 5974
 322 hrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlyGlyLeuSer 338
 :|||:.....
 5975 CTGAGAGTAATAGTGAGCCATTCTATATACGCGAGCGGAGATGTAGT 6024
 339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 355
 :|||:.....
 6025 CTTTCAGCAGACGAGGAGCATTTACCTCAATGGAATGCCATTGTTC 6074
 355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSera 371
 :|||:.....
 6075 AACACACACCAAACTACGAAAGAAATTCATATGACATAGACTACTG 6124
 371 lAlysMetThrAlaLeuArgSerSerAlaIleGlyArgAlaIleThrPhe 387
 :|||:.....
 6125 CAAGATACCAATTTACGTCATATCTGGCATACGATCTTTTCTTAC 6174
 388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLys 404
 :|||:.....
 6175 GATCCGATTAATGCTAATACGCTGCGGATTCCTACAGATTAATAATCT 6224
 404 lAsnGluThrProAlaAspSerAlaLeuGlnThrGlyAsnIleLeuP 421
 :|||:.....
 6225 CAATTAAGCTGATGACAGATATAGTACAGATTATAGGCGATGTGTT 6274
 421 heThrGlyGlyLysLeuSerGlnThrGluAlaAlaAspSerLysAsnLeu 437
 :|||:.....
 6275 TTCTGTGGAAGAACTCTCTGTAAGATGACGAAAGTTGCGAGAACCTC 6324
 438 ThrSerLysLeuGlnProValThrLeuSerGlyGlyThrLeuSerLe 454
 :|||:.....
 6325 ACTTCTACGCTGCAAGCAGCCTTAACCTTAACCTCAGAAATTTAGTACT 6374
 454 uLysHisGlyValThrLeuGlnThrGlnAlaPheThrGlnGlnAlaAsp 471
 :|||:.....
 6375 TTAACGAGGTGCTCACTCGATACGAAAGCTTACTACAGCCCGGCTT 6424
 471 eArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
 :|||:.....
 6425 CCTCTGTATATATGATGCGGACACAACGTTAAGAAGCAGTACAGAGAG 6474
 487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
 :|||:.....
 6475 GTCACTTTAAAGGCTTTCCATTCTGTACCTCTTACGCGAGGCTAA 6524
 503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSer 520
 :|||:.....
 6525 GAAAGTTGTATTTCTCTCTGACGACAGTAAATAATGTAACCTTACTG 6574

	7460	AGAAAGCTACAGAAGAAGATCTTTGATGACACCAACCCTTTCACATT	7509
	829	AlaLeuProIleGlyIleArgPheAspIysSeluSerAspScylaspaI	845
	7510	TCTTTCCTCCTAAGGGGTGAAGTTTGAGAAAGTTCCTGATTGTATGACTT	7559
	845	aThrIAsnLeuThrLeuGlyTYrThValAspLeuValArgSerAsp	862
	7560	TTCATTGATCTGACTTATCTCATGTTCCCGACTTATCCGAATGATC	7609
	862	IAspSpYThrThrThrIleuArgIleSerGlyAspSerTPlySThrPhe	878
	7610	CMAATGCACCTACACACTCTATATCACGCCGACCTCTTGGAACCTTAT	7659
	879	GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgIlaGlyAsnHisph	895
	7660	GCCAATAACTTAGCACACAGACGCTTGCAAGTGCGTCCAGCAGCTACTA	7709
	895	eCySPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuA	912
	7710	CGCTCTCTCTCCTATGTTGAAGTCTGGCCAGCTTGTCTTGAAGTTC	7759
	912	rGgLYSerSerArqAsnTyAsnValAspLeuGlyAlaTySyrGlnPhe	928
	7760	GTGATATCCACAGCATTTATATGTAATCTTGCGGTAAAGTTCACATTC	7809
seq_name:	gb_baz2.CPN133034		
seq_documentation_block:			
LOCUS	CPN133034	17280 bp	DNA
DEFINITION	Chlamydia pneumoniae GGAI-protein family cluster A, genes omp12 (partial), strain omp10, omp5, omp4, omp3 and omp14 (partial) and ORF1, strain VR1310.	BCT	24-FEB-1999
ACCESSION	AJ133034		
VERSION	AJ133034.1	GI:4455881	
KEYWORDS	hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13 gene; omp14 gene; omp5 gene; ORF1; outer membrane protein; outer membrane protein 11; outer membrane protein 12; outer membrane protein 14; outer membrane protein 4; outer membrane protein 5.		
SOURCE	Chlamydophila pneumoniae.		
ORGANISM	Chlamydophila pneumoniae.		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
AUTHORS	Boesen,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-JAN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, DK-8000 Aarhus, DENMARK		
REFERENCE	2 (bases 1 to 17280)		
AUTHORS	Hjerno,K., Boesen,T., Daugaard,L., Knudsen,K., Madsen,A., Christensen,G., and Birkegaard,S.		
TITLE	Chlamydia proteins containing the GGAI-repeat belong to a subfamily of autoantigenic pathogenicity factors		
JOURNAL	unpublished		
FEATURES	Location/Qualifiers		
source	1..17280		
	/organism="Chlamydophila pneumoniae"		
	/strain="VR1310"		
	/db_xref="taxon:83558"		
	1..841		
	/gene="omp12"		
gene	<1..841		
CDS	/gene="omp12"		
	/note="putative"		
	/codon_start=2		
	/translation="outer membrane protein 12"		
	/protein_id="CAB37067.1"		
	/db_xref="GI:4455882"		
	/translation="EDNIRRHNSGSVALGITATPAADQITFAFCQLFARDNRHHTG KNDHDTGASLVTHRRRTFDIANTRCKRPAAPVSRISQDIDCD		

FNEHMTAYTDNDILKSGSRNAPFCADQGLASLPYISVPILKKEEYFVVOYITPAQOOS
 YRHATGYTADPANNSELSINVEIPILGPIGFEEDBSSEGTQDLYLTMLVDFVYRRNPKHOS
 LILASDANMMAVIGTILANOCGSYVAANHFOVYNPHEITFGQFAFEVYSSSRNNTNLGSKR
 FCFP
 885. .968
 /gene="omp11"
 885. .3677
 /gene="omp11"
 885. .3677
 /gene="omp11"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein 11"
 /protein_id="CA937068.1"
 /db_xref="GI:4435883"
 /translation="MKPIRLKLLISLTVPILLSIATYAGDASLSPDSDFGAGGST
 FTPSPSTDANGINVTLSGVNYINDACKGALIGCCCTETTGDTLTGTGYSFENPVDY
 AGSNAAGASTATKARALVETGFENGLFIAPGATVYASKSTLSAGALNTLFEF
 SONSNEENNGGALITRKTLISGINSITPTSPNSKRKGGAYISSAASTISGNTGOL
 VPMNKKEGTGAGLFPASSITGNSLFPESCNTADAKGKGLAYSCKEPTPEPTD
 GSKSLTPFAENSSVTGCGALICANGDLDSAAQPTLPBNNCCNTAAGKGAIALADSGS
 LLSAGADGDTLPDNPNSPLDYSFTVSEKILYSSSKATINLEAAAGOSITFPDPAAS
 TGLSDAVLTITNPDPNSPLDYSFTVSEKILYSSSKATINLEAAAGOSITFPDPAAS
 ALKONVELDVNGEFLPOTEGSLTLMQPGTKLKADTEISLTKLVLDLSALEGNKSVSIE
 AGAKRTITLSPLEVFODSSGNEFESHITINQAFETQPLVETATVATASDIYIDALLTSPV
 OTPEPHGYOGHMTATWADISTAKSGTMMVTGTVNPNERASVYPPQSLWASTEDIR
 TLOOIMNSOANSITIOOGGLASGTANFPHKDKSGTQAPRRHSYCIYVGSAGADPESN
 IFSAVFCOLGSKMDDELYENTSHNVLASTYLYQHARFLOGGLPMPFSGITLMDLDIPL
 ILNADYSTSTTKNDMDRITRYPEAGDGLVNSNGALELGGSLATLPLPEAFPGGCTP
 FLKROAVSTRQONKRESGARFEDDGLVNSNGIPGLNIEISDEKNNEFISLAYI
 GDVYRKPNRSTSLMWSGASWTSICNKLARQAFLASGSHLTSHPVELSGEAYEELR
 GSAHYIVVDCGLRSTF"

gene
CDS

3827..6613
/gene="omp10"
3827..6613
/gene="omp10"
/note="putative"
/codon_start=1
/translation_table=11
/product="outer membrane protein"
/protein_id="CAB37069.1"
/db_xref="gi:4455884"
/translation="MKSSLLHMFLLISSLALPLSLNFEASFAAWEINLGPITSPFGPGG
YTPRATINADGTIYNLIGDYSITNAGSPPTLPLSCFEETGNLSFGCHGYPLQINNI
DAGANCTTNTPANKLLISFGSGFSYLSLIQTNNATTGTGATISTGCSIQSNVSCYFPG
NFSNNGALGSSSISLSLNPFLPFAKRAKQKGALVSTGGITITNLTNLSASENDEE
AANNNGALYTEKSSFSISNNKAISFIPNNKRAATGATGAIYCSSTAKPVLTLEEDNE
LGEINFTAITSGCAIYDNLVLSGGLPLFNFNNSAIDPAALGGIHALADSGSLSLSA
LGDITFGFNKTVYKGAASSQTTNRSINIGNNNAIKYDLRASQGNITFYDPIITTSY
AALSNALNLNGPDLGNRPATOGITVFSESEKLSAEAAEDNLKSTIQDPLTLAGGSL
LKSQYTLVAKSFSPGSGTLLMDAGTLETADGITNNLVNLSLKEETKATLTKATQ
ASQYITKSSLSLVPSCGNVEDVSMNRPQVFSCLITLADDPANIHITLADPLENNK
PIHMYQGNMALSMOEDPATKSKATLTWTKGNPNPERGRTIVANTLNSFEVDSVS
IIOVLATYVROSOERGIWCEGISMEFFKSTKINQGRHISAGYVCAATTLTASDML
ITAACTQFCGKDRDHFINKNRASAYAAHLQHLNLSPSLRLTPESSEQVLD
AOAISTIKDSEKMTYITQAPKGESSWVNGDGLAEASSLPTALSHGLFAPYFPFLIK
VAQSYIHDDSFERNETTLVRSFDSGLDINVSVPIGITFEFRSERNERASYEATVYVAD
VYRKPDCTTALLINNTSMKTTGTNLSPQAGIRAGIEYAFSPNLEVTNLSMEIRSSS
SRSYADLGKRFQ"

3827..3907
/gene="omp10"
complement(6954..8060)

sig_peptide

CDS

gene
CDS

sig_peptide

sig_peptide

gene	11535. .14321
	/gene="omp4"
CDS	11535. .14321

sig_peptide
gene
CDS

LMIHNLWKKPEDEPLFSLHYHCGRDYOAEFNSRAVDIIEPLFPAKTVDYDSHYHWYMA
 DKSYSGACGEHPLATYISGKERHNCGLISSNHCIEEDRQAVVSRITYEANSTIPLERLSLS
 OCGFASGEGHDHRIEEDQWKAIVAEINAEIKD" complement(18586)
 complement(18586)
 /gene="omp5"
 complement(8586, .11372)
 /gene="omp5"
 /codon_start=1
 /product="outer membrane protein 5"
 /protein_id="CAB37071.1"
 /db_xref="GI:4455867"
 /translation="MKQSFNSVLSTLACTSCSYFAATAENIGSPDSFEDGSTNTG
 TYTPNNTGIDYLTLDITLQNLGSPALATKRCFSDPTTESIFAGKGLSYFLILKIK
 SAEGALSVTTNDMLSTLGESPLATVPSGASPSVITPVGSGAGCGDITFDNNGIL
 KQDVEENGGAAITRNKLSINNGSTISFEENKSSAIGKRGKAGLACATGYDITNNAP
 LFDSNIHAAAGALINSNGCTITGNTSLVYESNSVTATNGAGALSADAVTISGNQ
 VTESNQNVAAGALIVAKKLTLAGSGGGGGSIFESNNIVQGTAGNGALSILAAEGCS
 VTESQNLNFAONAVTAPOTTKANEDISGTAKINLBAVLSHSLFFPDPTNTATNA
 DSDTDLINRKADNVAATDVGSIYFSGEKISDEDEAVANLISHTLQPVTLAGNLVIL
 KRGVLDKRGTPGTAGSSVIMADGDTLTKATEEVTLTGISIPVDSIEGKKVIVASNS
 ASKNALISGPLIIDNONGMANEHNDLGKQDSEFVQLSALGTTATTDVPAVPTVAT
 HGYGTGTGATVVDITASTRTKATATLANITELPNEROGPLVNSLMSGSFSDIO
 IOGVETALASLSDCRGMAAGVANEIDKDKGKERKRYHKSGGYVIGAATGCTCENNI
 ISFACQIFGSDKDFPLVAKHNTVTAACAFIOHITECSDICLDDPGSSHAPLILG
 LEGQALVSHENDIKTYTAVPEVKGSGNGMNAINLGAASSISPEYLHCPTPTVAT
 KLVITYIHDSPESEKIEGSGSDSNFNLSTLPGVKPEKFSKCNDSFYDLISVPLP
 LIRNDKQTTALITSGASWETIYANNLRAQLQVRAASHAFSPMEVFLQGFVFERGS
 SRHYVNDLGGKFGQ" complement(11313, .11372)
 complement(11313, .11372)

11535..14321
/gene="omp4"
11535..14321
/gene="omp4"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CA837072.1"
/db_xref="GI:4455887"
/translation="MKRSIPWLYSSVLSASCHLQSLANEELLSPDSENGINDSGT
LPKTSIATYKSLTGVGFYFEQGGKPLPSDSCFEKTTQNLVFLDNGSLHFLG
GAAPTATIKVLTSGFSLSPSPSTVYVLTSGAGVNELENTKGLVADNTHF
TADGAGLIGASFEFLTGTSGALSPNMSSTPKGALITVYTGARLANNQVRLSIA
TSGAGIDEEGSEIISLNKFTPEEGNAKTGGAICVTKRAGSEPLISNNKLIPAS
VAETSGGGIAHKKALTLSSGGFTPELRNNVSAAPAKNNTIFFDYPTISEGSSV
VRLNTLTGTGSDTPKRNAINSGNAKFTELRAKNNTIFFDYPTISEGSSVLT
GSAGLALVQSTLTLPSEETLPADKLVADMLKSSFTQPVSLSGKLLLOKQVLEST
FSQEGSLSDGSDSTLTSTAGSITTTNLGINDVSLGLQVSLPAKAGSNKYVFG
LNLIDENIETISMEFHDDFLSLKITVDADVTVDISSLIPVAPEDPNSEYFOG
OMNVWMTDTLTATKATATWTKTGVEPSERKSAVCTWLVGVSAGSNHPDLS
APAMHWKGFEWVSMTNELKTKGDEBNKGRFHTSGGVYVGGASNPKDLPFAC
FARDKDCIANNRSRTGTGLFPRKSHITLOPNVLELGRKPSFSAIEKFPREPLA
DYVQSFSDSDNRMTHTSLPSEGSNSNECINAGYIGLDLPVLSNPHRLKFTIP
KEMVYVSQNSFEPSDDSGKESTIGRLNLSIPVGAKFQGGIDGSVYTDLSGSG
VRYNRPDSTALVLSVPSMKIRGNLSRQAFLLRGSSNNVYVNSNCELFGHAMELRG
SVRYNPDVDTKLR"

NNHAYHDVNSNGAIAIAPGSGISISVKSGLDFKNTASODGNTIHSIHLOGCAOF
 KNLRAVSESGVYFDPIKSHSESHKITDLVINAPSEKREYEGTISFSGLCDHDEVCAR
 NLSTFILODVTLAGTSLSDVLTQLSHFQKOEASTLTMSGTTLGSGARVONLH
 ILIEDIDNEVPVRAIRAEKDKDALVSLERKLVAFEAWSYDPROFEARTIPLERIGP
 SPDSLLIGETTLERTQVTTENDAVGFMSLSNEERPSLDXDRITTPKTKIVFLTWNP
 EITSIP"

sig_peptide
 16494..16553
 /gene="omp14"
 16494..17280
 /gene="omp14"
 16494..>17280

CDS

alignment_scores:
 Quality: 2058.00 Length: 950
 Ratio: 2.831 Gaps: 20
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:

US-09-428-122-2 x CPN133034/rev ..

Align seg 1/1 to reverse of: CPN133034 from: 1 to: 17280

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePh 16
 |||||
 11372 ATGAAATCGCAATTTCTCGTTAGTCTCTTCGCACATTCGCATGTT 11323
 16 e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSer 30
 |
 11322 TACTAGTTGTTCCACTGTTTTCGCTGCACTGCTGAATAATATGCCCC 11273
 30 eRLAserPheAspGlyAsnLysAsn...GlyAsnPheSerValArgGlu 45
 |||||
 11272 CTGATGAGCTTTGACGAGACTACACACAGCAGCTATACCTCTTAA 11223
 46 SerGluGluAspAlaGlyThrThrTyLeuPheLysGlyAsnValThrLe 62
 |||||
 11222 ACG.....ACTACTGGAATAGACTATACCTGACAGAGATATACTCT 11179
 62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
 |||||
 11178 GCMAAACCTT...GGGAGATTGGCGACCTTTACGAAAGGTTGTTTCTG 11132
 79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
 |||||
 11131 ACGCTACGCAATCTTTAGCTTTGCGGTAAAGGGGACTCTCTTCTTT 11082
 96 GluThrValAspAlaGlyThrValAlaIleAlaValAsnSerVal 112
 |||||
 11081 TTAATATTAAGCT...ACTGCTGAAGCGCGACACTT...TCTGTTAC 11038
 112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIle 129
 |||||
 11037 AACTGATAAAATCTGCTGCTACAGAGATTTTGAGTCTTACTTCTTAG 10988
 129 lAserProGlySerSerIleThrThr....GlyLysGlyAlaValSer 143
 |||||
 10987 CGGCCCATCATCGTAATCACAAACCCCTCAGAAAGGTGACAGTTAA 10938
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
 |||||
 10937 TGTGGA...GGGAGCTTACATTGATTAACAAATGGAACATATTATTAA 10891
 160 rLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeu 177
 |||||
 10890 ACAGATTTACTGTGAGAAATGGCGAGCATTTCTACCAAGAACCTTT 10841
 177 eRLeuthrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
 |||||
 10840 CTTTGAAGAACGACGAGGATCTTTCTTGAAGGAAATTAATGAGC 10791
 194LysLysGlyAlaIleGlnThrSerAspAlaLeuThrI 207
 |||||
 10790 GCAACAGGAGAAAGGTGGGCTATTGTGCTACTGTTACTGTATGATAT 10741

207 eRLrGlyAsnGluGlyAlaIleSerPheSerAspAsnThrSerSerAsp 224
 |||||
 10740 TACAATTAATACGGCTCTACCTCTTCTCGAACAATATTCGTGAAGCTG 10691
 224 eRLGlyAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
 |||||
 10690 CAGCTGAGCTATTAATACACAGAACTGTAACATTAACAGGAATACG 10641
 241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThr 257
 |||||
 10640 TCTCTGTATTCTTGAATAATAGTGTGACA.....GCGACCCG 10603
 257 rGlyAspMetSerGlyAlaIleGlyAlaIleThrLysThrSerThrAsp 274
 |||||
 10602 AGCA.....AATGGAGAGCTCTT.....TCTGAGAGTG 10574
 274 hRLysValThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnThr 290
 |||||
 10573 CCGATGTACCAATATCTGGGAATCAGAGTGAACCTTCTCAGAAACCA 10524
 291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeuAla 307
 |||||
 10523 GCTGATGCTAATGCGGAGCCATTATGCTAAGAACCTTACACTGCTTC 10474
 307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLT 322
 |||||
 10473 CGGGGGGGGGGGGCGATCTCTTCTTCTACAAATATAGCCAGGATACA 10424
 322 hRLAProLysGlyAlaIleAlaIleGluAspSerGlyLeuLeuSer 338
 |||||
 10423 CTGCAGGTATATGTGGAGCCATTCTATCTGCGACGAGAGTGTAGT 10374
 339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
 |||||
 10373 CTTTCAGCGAAGACGAGGACATTACCTTCAATGGCAATGCCATTGTCG 10324
 355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSer 371
 |||||
 10323 AACTACACACAACTACAAAGAAATCTATGACATAGATGATCTAG 10274
 371 lAlaSerThrAlaLeuArgSerAlaAlaGlyArgAlaIleThrPheThr 387
 |||||
 10273 CAAGATACGCAATTTACGTCGAATATCGGCGATAGCATCTTTCTAC 10224
 388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLys 404
 |||||
 10223 GATCCGATTACTGCTAATACGGCTCGGATTTCTACAGATCTTTAACT 10174
 404 lAsnGluThrProAlaAspSerAlaLeuGlnThrThrGlyAsnIleLe 421
 |||||
 10173 CAATAAGCGTATGACGAGTAATAGACAGATTATAGTGGTGAATGTT 10124
 421 hethrGlyLysLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeu 437
 |||||
 10123 TTTCTGGTGAAAGCTCTCTGAAAGATGAAGCAAGTTGCAACACCACTC 10074
 438 ThrSerLysLeuLeuGlnProValThrLeuSerGlyLysThrLeuSer 454
 |||||
 10073 ACTTTACGCTTAACACAGCTGTAACTTAACCTACGACGAAATTTAGTCT 10024
 454 uLysLysGlyValThrLeuGlnThrGluAlaPheThrGlnAlaAsp 471
 |||||
 10023 TAAAGTGGTGCTACCTCGATACGAAGAGCTTACTCAGACCGGGGTT 9974
 471 eRLArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
 |||||
 9973 CCTGTGATTAATGATGCGGCGACACAGTTTAAAGCAATACAGAGAG 9924
 487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
 |||||
 9923 GTACCTTTAACAGTCTTCCATTCTCTGATGCTTTAGCGGAGGTTAA 9874

seq_name: gb_bai:AP002546	seq_documentation_block:	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	gene	CDS																																																																																																																																																																																																																																																																																													
AP002546	300600 bp	DNA	Chlamydomophila pneumoniae genomic DNA, complete sequence, section 2/4	AB002546	AB033782	AB033783	AB033784	AB033785	AB033801	AB033802	AB033803	AB033804	AB033805	AB033806	AB033807	AB033808	AB033809	AB033810	AB033811	AB033812	AB033813	AB033814	AB033815	AB036079	AB036080	AB036081	AB036082	AB038348	AB038349	AB038350	AB038351	AB038352	AB038353	AB038354	AB038355	AB038356	AB038357	AB038358	AB038359	AB038360	AB038361	AB038362	AB038363	AB038364	AB038365	AB038366	AB038367	AB038368	AB038369	AB038370	AB038371	AB038372	AB038373	AB038374	AB038375	AB038376	AB038377	AB038378	AB038379	AB038380	AB038381	AB038382	AB038383	AB038384	AB038385	AB038386	AB038387	AB038388	AB038389	AB038390	AB038391	AB038392	AB038393	AB038394	AB038395	AB038396	AB038397	AB038398	AB038399	AB038400	AB038401	AB038402	AB038403	AB038404	AB038405	AB038406	AB038407	AB038408	AB038409	AB038410	AB038411	AB038412	AB038413	AB038414	AB038415	AB038416	AB038417	AB038418	AB038419	AB038420	AB038421	AB038422	AB038423	AB038424	AB038425	AB038426	AB038427	AB038428	AB038429	AB038430	AB038431	AB038432	AB038433	AB038434	AB038435	AB038436	AB038437	AB038438	AB038439	AB038440	AB038441	AB038442	AB038443	AB038444	AB038445	AB038446	AB038447	AB038448	AB038449	AB038450	AB038451	AB038452	AB038453	AB038454	AB038455	AB038456	AB038457	AB038458	AB038459	AB038460	AB038461	AB038462	AB038463	AB038464	AB038465	AB038466	AB038467	AB038468	AB038469	AB038470	AB038471	AB038472	AB038473	AB038474	AB038475	AB038476	AB038477	AB038478	AB038479	AB038480	AB038481	AB038482	AB038483	AB038484	AB038485	AB038486	AB038487	AB038488	AB038489	AB038490	AB038491	AB038492	AB038493	AB038494	AB038495	AB038496	AB038497	AB038498	AB038499	AB038500	AB038501	AB038502	AB038503	AB038504	AB038505	AB038506	AB038507	AB038508	AB038509	AB038510	AB038511	AB038512	AB038513	AB038514	AB038515	AB038516	AB038517	AB038518	AB038519	AB038520	AB038521	AB038522	AB038523	AB038524	AB038525	AB038526	AB038527	AB038528	AB038529	AB038530	AB038531	AB038532	AB038533	AB038534	AB038535	AB038536	AB038537	AB038538	AB038539	AB038540	AB038541	AB038542	AB038543	AB038544	AB038545	AB038546	AB038547	AB038548	AB038549	AB038550	AB038551	AB038552	AB038553	AB038554	AB038555	AB038556	AB038557	AB038558	AB038559	AB038560	AB038561	AB038562	AB038563	AB038564	AB038565	AB038566	AB038567	AB038568	AB038569	AB038570	AB038571	AB038572	AB038573	AB038574	AB038575	AB038576	AB038577	AB038578	AB038579	AB038580	AB038581	AB038582	AB038583	AB038584	AB038585	AB038586	AB038587	AB038588	AB038589	AB038590	AB038591	AB038592	AB038593	AB038594	AB038595	AB038596	AB038597	AB038598	AB038599	AB038600	AB038601	AB038602	AB038603	AB038604	AB038605	AB038606	AB038607	AB038608	AB038609	AB038610	AB038611	AB038612	AB038613	AB038614	AB038615	AB038616	AB038617	AB038618	AB038619	AB038620	AB038

/gene="CpJ0267"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98477.1"
/db_xref="GI:8978641"
/translation="MSIMSLINKTNALLNPPEPAVCLNAMPKYINODRKTFACVYTL
/translatingVILLANGSPGLSVSTIIIGSTIGTALFIIGLVKLIKSLIAMI
OYKOEYEVKQKYEFSIPKNDNHKLLSCLPSPLDIESPSEASTPYSKRIACSG
VAIVGLVTLIGAVSVFPCYGLQIALVGVACIACIATLFFVGLAGLRHSLIAOGIM
YIYLYIYSSALREERNEIVKDRNEINITYLIEECROKREKALLE"
887..1273
/gene="CpJ0268"
887..1273
/gene="CpJ0268"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98478.1"
/db_xref="GI:8978642"
/translation="MSOCSSSTSTWEMKSFVPMKNPTPLSPISPEDETLIAYEP
FVLPTDPENAOANPGTSTPVGNGIDDLNLLGQPNQNNANNPGTSGSNPTSLPA
PRLPTEFENSGEEOGSONNEDLIG"
complement(1425..2402)
/gene="CpJ0269"
complement(1425..2402)
/gene="CpJ0269"
complement(1425..2402)
/codon_start=1
/transl_table=11
/product="dipeptidase"
/protein_id="BA98479.1"
/db_xref="GI:8978643"
/translation="MTIDMHCDSLSPHFRCRKPDAVRCSPQELLSGVROQCAIFV
HSRGPNDCKONSLEFSPNCPYDGLISYEENSGSSOKSLISRSIENASLGD
DIAPGTLIAKLILHJTGQPLAIGIVMKGNRFGCGTEAPKRLSNDCKVLLDITVEL
GYPIDSHCSDKLADIDIDYTDKLPINAVISHNFPSVDHRNLDLDAHKELIVR
KVYGLNLYRSTVGDSDGLEKHVLAENLGLSSIVGSDGFYANEDENEFNECCSS
AEAPVLNQLHRIFSKGAESILSSRAEKFLLQYIVQVNPVKITDVXL"
complement(2417..3277)
/gene="ywlC"
complement(2417..3277)
/codon_start=1
/transl_table=11
/product="SuA5 superfamily-related protein"
/protein_id="BA98480.1"
/db_xref="GI:8978644"
/translation="MPDKAQTSTSLPEVMSAHQKIVALPTDVVGFVLSIASEA
EERLYALDRPSKAFALYVNSIEDIENISGPISTPAKKLAQLFPGAITLVKHNRP
RPPKTLARIVDSVREIVDHGCTLIGTSANLSEFSAITAEITFADFDHLCIF
DGPCSHGLESTVVASDPLIYREGLSRSVTEINAGTEAKIFHRTSHAFSKHIXITY
KNQDLVSFLSGSDFGKGVCEHPRKMFYRLREALKKTPTSIVFIYDINTSYPEL
FPFLSPYIE"
3586..4317
/gene="CpJ0271"
3586..4317
/gene="CpJ0271"
/codon_start=1
/transl_table=11
/product="lysophospholipase esterase"
/protein_id="BA98481.1"
/db_xref="GI:8978645"
/translation="WTDYSFRRKIGNIEALECPGNPDPIILCHGVGSLADNLTEFF
PSICSFKLRPTWEPNGILPLENDPFGSRACFPNVLLOELSRILYANGVGNLOEKY
DELFDVLETKALEELILNLRPNYNEIIIGGFSQGLIATLHVLISQNYAALAE
AGARLFGNGWEGGKCGAOVPELOSHGEDILPYHLGAHLNDLLIKLNQPFSTFG
GHEIPSVFQCKQWOTVPMWIDPARG"
complement(4289..5185)
/gene="dnaX_2"
complement(4289..5185)
/gene="dnaX_2"
complement(4289..5185)
/codon_start=1
/transl_table=11

/product="DNA polymerase III gamma and tau"
/protein_id="BA98482.1"
/db_xref="GI:8978646"
/translation="MHLSEENQGWELARKVYHOEVPAILLHGFTLPVLQDKAEOLA
SEILLSGSPGSEHKVSOKIHPIDIVQFPEGGRHSDIDPGIKOYIISPEFANYKI
YIIEADRMTLAAISAFKVEEPPKHAVIILLTAKVORLPKTIISRSLSIFIRGK
ILCSKETSTFVRYAOCCEIPVTEYSOIIKESSEEDKQVLRKQVPMFVLELIRDY
TLNIGKASALNYEBHKHEILQLEPLPLDKVLLIVESCSNLNNSASVLEWVAIO
LVSIOYKRELEVSPGODSN"
complement(5176..5796)
/gene="tdk"
complement(5176..5796)
/gene="tdk"
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="BA98483.1"
/db_xref="GI:8978647"
/translation="MFIYIEGEGSGKSLAKALGDLQVADRKVILLTREPCCLTIGE
RLRDLLEPPLHLEISRCCELPFISGRAOHOIPIVIALDGYIVTEFRRHDSITYO
GIACIGADPADICSVYGPPTPLPFPVLLIDIPADIGLQKRRHQVDFEKKPLS
YHNIRIEGFLSLASADSRIVLVLDRESLASLIDKVMHLTQLGCT"
complement(5801..8305)
/gene="gyrA_1"
complement(5801..8305)
/gene="gyrA_1"
complement(5801..8305)
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BA98484.1"
/db_xref="GI:8978648"
/translation="MFKNDETIYVKNLEEMKESLYRSMSVIISRALPIDRLGKPS
QORVLVYMKOISLSPKCAKHKRCAKICGDTSGDHYRHESTYPIVYADNANRYPL
VDGKNGKSDSGDPPARMKRTERTALTIHSANIMEDLDKDVIVDPNYDKHEPVP
SKFPNLCONSGSIAVGMATNIPHNIGELIEATLTLAPQASVEDIILQVMPGPPT
TGIIICGSEGIIRSTYTTGRGKIKVRAHLEKGESESEIINLYKFTDVQVFGAM
IANLVNKTLAGISIDVDESDKDGIIVALEIKGESESEIINLYKFTDVQVFGAM
LADKNPRTMSIIRMSISAMIRHREKVIYRRTRETLKATRAHVLEGYKALSCDA
LVKTIRESGNKHAKEKRIIESFGTEQALAIETRLYOLGIEAEIKOYEBELNK
IAYKQVLSDEGIVKDIIRNELODLKHKHVARITIEFPADIDREDIITNBSYII
TISGDDYKRPKPVKVEKQRRGGHGVTFDMKKGAGFLKAVYSAFTKDYLLITNFGQ
CYWLKWOLEGEGERAKRPIINLEGIIRGEELAAILNKFNPDNGFLFLATRGVY
KVSYLDAFSNPRKGIARLIDEGDELIAACHIVISDEKVMFLTHGMAYFPHEVYR
PMGRTAGVGRVSLKNEDKEDVSCQIVTENQSVLYVGDGFGKSLVEDERETNRCGV
GVRSLIINENGNVLAIPVTDHDSIILMSOGQAIRINNQDVAVMGSRSTQGVRLVHL
KEGDALYSMEKLSNENDEVLISGSEECSDTYSLR"
complement(8321..10738)
/gene="gyrB_1"
complement(8321..10738)
/gene="gyrB_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BA98485.1"
/db_xref="GI:8978649"
/translation="MDPEKKNYDASAITVLEGLQAVREPMYIGDTGITGLHNLVE
LVVNSIDEMAGYCSRIDVRILEDGIVAVDNRGPIEYHERESAKOGREYALVY
ITVLHAGGCFDKDSYKSGSLHGVGSVNAISGKIVATVFKKQKQOMFSGRGIPT
PLQYVYSDRQGEIIVYBPDKTIESCTDEPQALIRLRELATLNGITIVPEDDDV
SEPKVTFYFGIGOSFYSYINOKKESLSEPIITICTRGDDSEIEFEALQONSGIS
ELVYSTANNPIRQGGTHLGFSTALTRYVINYIKAHNLKANRKLALTEGDIAREGLTA
VISEAVKPNPEQGTQOKLGNDSVSAQVWGEALTIFFEENPQIARIMVDVAFVAA
QAREAAKARELILRKLSALDSARLPGLKIDLCDEKDEKMYIVEDSGAGSAGKQGR
RRQAVIPIPIGKLIVNEKARLOKIFPNOEIGTIIAAGGIGADNENLSKLRYRRII
MTQADVGSIRLITLLTFPRHMTALINDCVYIAQPLKYVSKKDDFYITSEKMD
SYILMIGTNSLSLFTSTERELGALBESTINVIDVESINLEKKAIPSEFLEMY
KEBIGPLVLYLAPATGMOGRVLYSDEEKEALAOETKFKIILELYKAAVFIDIONQ
LKEIGYGLDISSVLIPOKNEIYVIGNEDSPSCNYSCTLEEVINYLNKNGRGIETQRYK
LGEMNADQMDWTMNEORTLIHVSJKDAVEADHIFTMLGEEVPPRREFIESHALSI
RINNLDT"
complement(10742..11089)
/gene="CpJ0276"
complement(10742..11089)

```
/gene="CPj0276"
/codon_start=1
/transl_table=11
/product="Ct191 hypothetical protein"
/protein_id="BA098486.1"
/db_xref="GI:8978650"
/translation="MFLKRRKGGGSOVONKRTASPIKAKHYLHNYLOELOKIMARP
HDAIDAMNOVFRDRIKGMSSQAIQFRDHIILVKNVNSLYALLKQIPONDLMISLYOVA
SHVOIREIOFLG"
complement(11353..11862)
/gene="CPj0277"
complement(11353..11862)
/gene="CPj0277"
/codon_start=1
/transl_table=11

alignment_scores:
Quality: 2058.00      Length: 950
Ratio: 2.831          Gaps: 20
Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:
US-09-428-122-2 x AP002546/rev ..
Align seg 1/1 to reverse of: AP002546 from: 1 to: 300600

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePh 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207760 ATGAAATCGCAATTTCTCGTGTAGCTCTCTCTCTGACATGCGATGTTT 207711
16 e.....ProLysSerMetIleAlaThrGluThrValLeuAspSers 30
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207710 TACAGTGTTCCTGCTGTTTCTGCTGCACTGCTGAAATATAGCCCT 207661
30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgGlu 45
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207660 CTGATACCTTTCAGCAAGTACTTAACACAGCAGCTTACTCTAAAAAT 207611
46 SerClnLysPheAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe 62
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207610 ACG.....ACTACTGGAATAGACTATCTCTGACAGAGATATAACTCT 207567
62 uGluAsnIleProGlyThrThrAlaIleThrLysSerCysPheAsn 79
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207566 GCAAAACCTT...GGGATTCGGCAGCCTTACACAGAGGTTTCTG 207520
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207519 ACACCTACGGAATCTTTAAGCTTTCGCGTAAGGGGTACTCTTTCTTT 207470
96 GlnThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerVa 112
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207469 TTAATATTTAAGTCT...AGTGTGAAGGCGCACACTT...TCTGTTAC 207426
112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207425 AACTGATATAAAATGTCGCTACACAGATTTTCAGCTTACTTCTTAG 207376
129 lAsnProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207375 CGGCCCATATCGGTATGATCAACCCCTCAGGAAAGGCGAGTAA 207326
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207325 TGTGGA...GGGATCTTACTTTGATATACAAATGAACTATTTATTTAA 207279
160 rLysAsnPheSerThrAspAsnGlyAlaAlaIleThrAlaLysThrLeuS 177
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207278 ACAAGATTACTGTGAGAAATGCGGAGCCATTTTACCAAGATCTTT 207229
177 eirLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207228 CTTTGAAGAACGACGCGATCTTTCTTTGAAGGAATATAATCGAGC 207179
```

```
194 .....LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrI 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207178 GCAACAGCAGAAAAAGGTGGGCTATTTGCTACGTGCTACGATAT 207129
207 eThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSerAspS 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207128 TACAAATAATACGGCTCTTACCTCTCTGAAACAAATTTGTAAGCTG 207079
224 eGlnAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAla 240
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207078 CAGGTGAGCTATTAATATAGCACAGAACTGTACATTAAGGGAATAGC 207029
241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThrTh 257
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207028 TCTCTGTATTTCTGAAATAGTGTGACA.....GCGACCCG 206991
257 rGlyAspMetSerGlyAlaIleCysAlaIleThrLysThrSerThrAsp 274
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206990 AGGA.....AATGGAGAGCTCTT.....TCTGGAGATG 206962
274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnAsnThr 290
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206961 CCGATGTTACCATATGTCGGGAATCAGAGTGTACTTCTCAGAAACCA 206912
291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeuAlaSe 307
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206911 GCTGTAGCTAATGCGGAGCCATTTATGCTAAGAACCTTACACTGCTTC 206862
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyL 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206861 CCGGGGGGGGGGGGATATCTCCTTTCTTACATATATAGTCCAAAGTACCA 206812
322 hrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeuSer 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206811 CTCACAGTATGCTGAGCCATTTCTATCTGACACTGAGAGTGTAG 206762
339 LeuSerLysAspSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206761 CTTTCACAGACAGACAGGAGCATTTACCTTCAATGGAAGGCCATTTGTC 206712
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSer 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206711 AACTACACCCAAACACTACGAAAGAAATTTCTATTTGACATAGATCTAC 206662
371 lAlysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleThrPheThr 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206661 CAAAGATCACAGAAATTTACGTCGAATATCTGGGCATTCGACATTTTCTAC 206612
388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206611 GATCCGATTTCTCTATATAGGCTCGGATTTCTACGATATCTTAAATCT 206562
404 lAsnGluThrProAlaAspSerAlaLeuGlnThrGlyAsnIleLeP 421
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206561 CAATTAAGGCTGATGACAGATATGATATGATATGAGGTCATTTGTT 206512
421 heThrGlyGluLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeu 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206511 TTCTGTGTAAGGCTCTGGAAGATGAACAAAGTTGAGCAACCTC 206462
438 ThrSerLysLeuGlnProValThrLeuSerGlyGlyThrLeuSerLe 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206461 ACTTCAAGCTGAGAGAGCCTGTACTTACCTACGACGGAATTTATATAC 206412
454 uLysAsnGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspS 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206411 TAAACGGGTGTCTGCTGATGAGAAAGCTTTACTGACCCGCGGT 206362
471 eirArgLeuGluMetAspValGlyThrThrLeuGlnProAla...AspThr 486
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206361 CTTCTGTATTATATGATGCGGACACAACGTTAAACCAAGTACAGAGAG 206312
```


[illegible]

Alignment_scores:	Quality: 2044.00	Length: 950
Ratio: 2.804	Gaps: 20	
Percent Similarity: 76.737	Percent Identity: 46.737	
<p>Alignment_block: US-09-428-122-2 x AE001628/rev ...</p>		
<p>Align seg 1/1 to reverse of: AE001628 from: 1 to: 10757</p>		
1	MeltySerSerPheProLysPheValPhe...SerThrPheAlaIlePhe	16
7650	ATGAATTCGCAATTTTCCTGGTAGTCTCTCTCGACATTCGCATGCTTT	7601
16	e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSer	30
7600	TACTAGTTGTTCCACTGTTTCTCTGCACTGCTGAATAATATAGCCCT	7551
30	erAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgLys	45
7550	CTGATAGCTTTCAGCGAAGTAACTAACAAGACAGCACTTACTCTCTAAAT	7501
46	SeGInGluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe	62
7500	ACG.....ACTACTGGAATAGCACTAATCTCTGACAGCAATATTAACCT	7457
62	uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn	79
7456	GCAAACCTT...GGGGATTCGGCAGCTTAAAGAAAGGTTGTTTCTG	7410
79	snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe	95
7409	ACCTTCACGGAATCTTTAAGCTTTGCGGGTAAGGGTACTACTTTCTTTT	7360
96	GlnThrValAspAlaGlyThrValAlaGlyAlaValAlaValAsnSerSer	112
7359	TTRAAATTAAGCT...AGTCGTGAAGCGCACACTT...TCTGTAC	7316
112	ValAlaPlySerSerThrThrPheIleGlyPheSerSerLeuSerPheIle	129
7315	AACTGATTAATAAACTCTCCCTACAGCAATTTTGAGCTTACTTTCTTAC	7266
129	IaSerProGlySerSerIleThrThr....GlyLysGlyAlaValSer	143
7265	CGGCCCATCATCGTAACTACCAACCCCTCAGCAAAAGGTGCACTTAA	7216
144	CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe	160
7215	TGNGGA...GGGGATCTTACATTGTAACAATGAACTATTTTATTAA	7169
160	LysAsnPheSerSerThrAspAsnGlyGlyAlaIleThrAlaLysThrLeu	177
7168	ACAAAGTACTGTGAGAAATATGGCGAGCACTTTCACCAAGATTTT	7119
177	erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer	193
7118	CTTGAAAAACACACGCGATTCATTCTTTGCAAGCAATTAATCAGC	7069
194LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrI	207
7068	GCAAACGAGAAAAAAGCTGGCGCTATTTCGACTACGTCTAGTAGAT	7019
207	erThrGlyAsnGlnGlyValSerPheSerAspAsnThrSerSerAsp	224
7018	TACAAATTAATACGGCTCCTAACCTCTTCTGCAACAATATTCGTAACCTG	6969
224	erGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAspAla	240
6968	CAGGTGACTATTAATATGACACAGCAAGCACTCAATTAACAGGAATACG	6919
241	LysValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThr	257
6918	TCTCTGTATTTCGTAATAATATGTGACA.....GCGAGCGC	6881

553 tValaIserSerThrAlaValThrProAspProLeuMetLysLysPheH 570
6002 TCGCAACAACTACAGATGTTCCAGCGGTTCCACATATACCACTCCCTACG 59533
570 1stYrG1yGrcInG1yThrTpg1yProIleValTyr..... 582
5952 ACHATGGGTATCAAGTACTCTGGGA...ATGACTGGGTGATGATAC 59068
583 ...GlyThG1yAlaSerThrThAlaThrPheAsnTprThrLysThG1 598
5905 GCAAGACACCCAAAGACTAAGACACGACGACATAGCTTGACCAATACAG 5856
598 YTYrIleProAsnProG1yArg1LeG1SerLeuValProAsnSerLeuT 615
5855 CTAACCTCGAATCCTGAGCGTCAGAGACCTTTACTCTTAATACCTT 58068
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGluThAla 631
5805 GGGGATCTTTTTCACACATCCAAAGCATTCAGAGTGCATAGAGAGAGT 5756
632 AsnGluG1yLeuG1yAspArg1aPheTyrG1yArg1yLeuSerAs 648
5755 GCTTGACCTCTTGTTCAGATGAGGCTTTGGCTGGGAGTCCGCA 57068
648 nPhePheHisLysAspSerThrLysThArgArgG1yPheArgHisLeuS 665
5705 TTTCTTAGATAAAGATAGAAAGGGAAGAAAGCAAAATACCGCATAAAT 5656
665 eTcG1yG1yTyrAlaIleG1yG1yAsnLeuHisThCysSerAspLysIle 681
5655 CGGGGAGATGTCATCGGAAGTGCAGCGCAACCTGTTCGAAACCTTA 56068
682 LeuSerAlaAlaPheCysG1yLeuPheG1yArgAspArgAspTyrPheVa 698
5605 ATTACCTTGCTTGGCAACTCTTGGTAGCGATTAAGATTCTTAGT 5556
698 1AlaLysAsnG1yThrValTyrG1yG1yThLeuTyrTyrGlnHisA 715
5555 CCGTAAATCATCTACTGATACCTATGCGAGGACCTTCATATCCAAAC 55070
715 nG1yThTyrIleSerLeuProCysLysLeuArgProCysSerSerLeuS 731
5506ATTACAAATGTAGTGGGTTCCATAGTGTCTCTTAGAT 5468
732 TyrValPro.....ThGluIleProValLeuPheSerG1yAsnLe 745
5467 AAACCTTCCTGGCTCTGGAGCATTAACCCCTGTTTAGAAGGCGAGCT 5418
745 uSerTyrHisThrAspAsnAspLeuLysThLysTyrThrThrTyrP 762
5417 CGCTTATAGCCACGTCATATGATCTGAAACAACAAGATATCTCGTATC 5368
762 rOthValLysG1ySerTrrPheLysAsnAspSerPheAlaLeuG1yPheG1y 778
5367 CTGAGGTGAAGAGTCTCTGGGGATAATAGCTTTAATCATGATGTGGGA 5318
779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluG1yTyrMe 795
5317 GCTTCTTCATTTCTTATCCGAATACCTCGCATTTTGTGATACCTATGC 5268
795 tProPheMetLysLeuGlnPheValTyrAlaHisG1yGluG1yPheLysG 812
5267 TCCATACATCAAACTGAATCTGACCTATATACGTCAGAGACAGCTTCCG 5218
812 LuG1yG1yThrGluAlaArg1ubPheG1ySerSerArgLeuValAsnLeu 828
5217 AGAAAGGTACAGAAAGAGATCTTTGATGACACACCAACCTTCATTAATTA 5168
829 AlaLeuProIleG1yLeuArgPheAspLysGluSerAspCysGlnAspAl 845
5167 TCTTTGGCTATAGGGGTCAATTTTGAGAAGTTCCTGATTTGATGACTT 5118
845 aThrTyrAsnLeuThrLeuG1yTyrThrValAspLeuValArgSerAsn 862

```

.....
5117 TTCTTAGATGACTTATTCCTATGCTCTGATCTTATCCGAATGATC 5068
862 rospcythrthrthleuarglleserlyaspserthrplsthrphe 878
11 |||||..... |||||.....
5067 CCNAATGACATACAGCACTGTATACGGAGCCTTTGGAAACTAT 5018
879 GlyThrAsnLeuAlaarglAlaLeuValleuArgAlaGlyAsnHisph 895
..... |||||.....
5017 GCCAATAACTAGACAGCAGCGCTTCAAGTGCCTCAGCAGCATCA 4968
895 ecyspheasSerasphegluaIapheserGlnpheSerPhegluleuA 912
..... |||||.....
4967 CGCCTTCTCTCCTATGTTTGAAGTGGCTGGCCACTTGTCTTGAACTTC 4918
912 rgllySerSerArgasnlyrAsnValaspleuGlyAlaIalstYrGlnPhe 928
|||||..... |||||.....
4917 GTGATCTCAGCGATTTATATAGTACATCTTGGGGTAAGTTCCAATTC 4868
seq_name: gp_pat:A81837

```

```

seq_documentation_block:
LOCUS      A81837          2757 bp      DNA
DEFINITION Sequence 11 from Patent WO9858953.
ACCESSION  A81837
VERSION    A81837.1 GI:6731869
KEYWORDS   .
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 2757)
AUTHORS     Madsen, A. and Birkelund, S.
TITLE        NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL      Patent: WO 9858953-A 30-DEC-1998;
             MADSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)
FEATURES
     source              1..2757
                        /organism="unidentified"
                        /db_xref="taxon:33644"
BASE COUNT   770 a 564 c 642 g 781 t
ORIGIN

```

```

alignment_scores:
      Quality: 2001.00      length: 953
      Ratio: 2.814      Gaps: 21
      Percent Similarity: 74.607      Percent Identity: 45.435

```

alignment_block:
US-09-428-122-2 x A81837 ..

Align seg 1/1 to: A81837 from: 1 to: 2757

```

1 MetlySerSerPheProlySphValPheSerThrPheAlaIlePhepr 17
|||||..... |||||.....
1 ATGAGATCGCTTTTCTTCTTATTAATCTTCATCTCTACGCTTTC 50
17 oleuSerMetIle.....AlaThrGluThrValLeuAspSer 30
|||||..... |||||.....
51 TCTCTTAATGAGTCTTCTGACATGCTGCCGATCTCACATTAGGGAGTC 100
30 erLaserPheaspGlyAsn...LysasnGlyAsnpheserValaArglu 45
|||||..... |||||.....
101 GTGACAGTTATATGATGATACAGCACACAGATTACTCTCTAAAGCG 150
46 SerGlnGluAspAla...GlyThrThrThrLeuPheLysGlyAsnValth 61
|||||..... |||||.....
151 GCAACTTCTGATGCTAGTGCAGACCTATATCTCGATGGGATGTCTC 200
61 rleuGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheA 78
|||||..... |||||.....
201 GATTAAGCAA...GCAGGAAACAACAGACCTTAACCAAGTTGTTT 247

```

```

78 snAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeu 94
|||||..... |||||.....
248 CTACACACTCAGCAAAATCTTACCTCTTAGGCAACGATTTCTCTTCA 297
95 PheGlnThrValasPheAlaGlyThrValaIleGlyAlaValasSerse 111
|||||..... |||||.....
298 TTTCACAAATATTTTTCGTCTACTGTTGCAAGTGTGTGTGTAGCATAC 347
111 rValValasPlySerThrThrPheIleGlyPheSerSerLeuSerPhe 128
|||||..... |||||.....
348 AGCAGCTCTGGGATTAACGAATCTTCAGAGATTCTCAACTCTCGGATGC 397
128 leuLaserProGlySerSerIleThrThrGlyLysGlyAlaValasCys 144
|||||..... |||||.....
398 TTGACAGCTCTTAG.....ACCACAGTAAGAGCCATTAATAAT 438
145 SerThrGlySerLeu.....SerLeuThrLysAsnValse 156
|||||..... |||||.....
439 ACCGATGCTCTGTGTGTGAGAGTATAGGAATCTTGACCAAAAT..... 483
156 rleuLeuPheSerLysAsnpheserThrAspAsnGlyAlaIleThrA 173
|||||..... |||||.....
484 .....GAAAATGCCCTCTAGTGAATAAGGGGAGGCATCAATA 520
173 lalysThrLeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerGlu 189
|||||..... |||||.....
521 CGAAGACTTGTCTTGTGACTGGAGTACCGGTTGTGACTCTTCTTGGC 570
190 AsnThrSerSerLysGlyGlyAlaIleGlnThrSerAspAlaLeuTh 206
|||||..... |||||.....
571 AATAGCTGTCGCAACAGAGGAGGAGATCTATGCTCTGGAGACTCT 620
206 rIleThrGlyAsnGlnGlyGlyAlaSerPheSerAspAsnThrSerA 223
|||||..... |||||.....
621 GATTTGTGAATGACAGCAATCTTGCAGTTCGAAACAACACTGCCACA 670
223 spSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsn 239
|||||..... |||||.....
671 CATCAGAGAGCGGATCTCTCTGGAAGGACCTTGATCTCCATATAC 720
240 AlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSerTh 256
|||||..... |||||.....
721 CAATAATATCTTTTCGATGCTGCAACGCACT..... 753
256 rThrGlyAspMetSerGlyGlyAlaIleCysAlaIyrlys.....Thrs 271
754 .....ACAAATGCGCGAGCTATTTGATTGATACAAAGCAGGGCGCA 793
271 erThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSer 287
|||||..... |||||.....
794 ACCCAGACCCCTATCTTGACTCTTTCAGAAATGAGAGCCTGCATTTTC 843
288 AsnAsnThrSerThrAlaGlyAlaIleIleThrValLysLysLeuG 304
|||||..... |||||.....
844 AATAACACAGCAGAAATAGTGAGGTGCGATTATACCAAAAATTCGT 893
304 uleuAlaSerGly...GlyLeuThrLeuPheSerArgAsnSerValasG 320
|||||..... |||||.....
894 GTTATCTCAGGACGAGGAGAGTGTATTATTCTAACCAACAAGCTCGCA 943
320 lylGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlylu 336
|||||..... |||||.....
944 ATGCTACT...CCTAAGAGAGGGCAATTCGATTCTAATCTCGAGAG 990
337 leuSerLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThyA 353
|||||..... |||||.....
991 ATTAGATTTCGACAGATCTCGCAATATCAATTTTCGAGGGAATAACT.. 1038
353 lThrSerThrThr.....ProGlyThrAsnArgSerSerIleAspL 367
|||||..... |||||.....
1039 .ACGACACTACAGAGAGTCGCGAGTGTGACCAAGAAATGCTATAGATC 1087
367 euGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIyArgAla 383

```



```

1088 TTGCATCGAATGCAAAATTTTAAATCTCCGAGGAGCTCGGGGAAATAA 1137
384 ILeryrherlyrasproliethrthGlySerSerThrValThrAs 400
1138 GTTATTTTCTATGATCCTATCAGC.....AGCTCAGAGAGCTACTGA 1178
400 pValLeuLyValAsnGlnThrProAlaAspSerAlaLeuGlnTythrG 417
1179 TAGCTCTCTTTGATATAAGCTGACGAGATCGAATAAATCACTATGAG 1228
417 LysnIleIlePheThrGlyLysLeuSerGlnThrGlnAlaAlaAsp 433
1229 GCTACATCGTTTCTCTCGAGAGAACTCTCAGAGAGAACTTAAGAA 1278
434 SerLysnLeuThrSerLysLeuLeuGlnProValThrLeuSerGly 450
1279 CCGACATGCTGAAGTCTACATTTACACAGGCTGTAGAGCTTGTGAGG 1328
450 yThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
1329 TGCCCTAGTATGTAAGATGAGTGAAGTCTAGTTCGAAATACATAAGCC 1378
467 InGlnAlaAspSerArgLeuGluMetCaspValGlyThrThrLeuGlu... 482
1379 AGGTCAGAGGATCGAAAGTCGTATGATGAGGAGGAGCTACTTTGAGGCA 1428
483 ProAlaAspSerThrIleAsnAsnLeuValIleAsnIleSerSerI 499
1429 AGCGCTGAGGGGGTCACTCTCAATGGCTTACCCATTATATAGATTCTT 1478
499 eAspGlyAlaLysLysAlaLysIleGlnThrLysAlaThrSerLysAsn 516
1479 AGATGGGACAAATAAAGCTATCATTAAGCGAGCGACGACAGTAAGAGT 1528
516 euthrLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTy 532
1529 TTGCCCTATCAGGGGCTATCATGCTGTAGATGCTCAGGGGAACTATTAT 1578
533 GluAsnHisSerLeuArgAsnProGlnSerTyArgIleLeuGlnLeu 549
1579 GAGCATCATATATCTCAGTCACAGCAGGCTTCTCTTATATAGAGCTTTC 1628
549 sAlaSerGlyThrValThrSerThrAlaValThrProAspProIleMe 566
1629 TCACACAGAGACGATGACTACTACAGATATCCCGCATACCCCAATTCTTA 1678
566 LysGluLysPheHisTyArgLysGlnGlyThrTyrGlyProIleValTyr 582
1679 ATACTACGAAATCAATGGGATACAGGAACT...GGAATTAATTTGTTGG 1725
583 .....GlyThrGlyAlaSerThrThrAlaThrPheAsnTyrThr 596
1726 GTCGACGATGCAATGCAAAACAAATAATGCTACTTAAGTGAAGCTTA 1775
596 sThrGlyTyrIleProAsnProGlnArgIleGlySerLeuValProAsn 613
1776 AACAGGATCAAGCCGAATCCAGAACGTCAGGACCTTTGGTTCCTAATA 1825
613 eLeuThrAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGlu 629
1826 GCGTGGGGGCTCTTTGTCGATGTCGCTCCATTCACAGCCTCAATGGAC 1875
630 ThrAlaAsnGluGlyLeuGlnGlyAspArgAlaPheTyrCysAlaGly 646
1876 CGGACACACAAAGTTCGTATCTCTGTCACAAATTTGTGGGTACAGAA 1925
646 userAsnPhePheHisLysAspSerThrLysThrArgGlyPheArgH 663
1926 CGCGGAGCTTTTGGCATGAGATCAGAAAGAAACCAACGTAATTTATCTC 1975
663 ILeuSerGlyGlyTyrValIleGlyLysLeuHisThrCysSerAsp 679

```

```

1976 ATTCTAGCCCGGCTTATGCTATTAGAGAGATTTCTTACGGCTTCTGAA 2025
680 LysIleLeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAsp 696
2026 AATTCTTTAATTGCTTTTGTGTCAGCTTTTGGCTACGACAAAGACCA 2075
696 rPheValAlaLysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTy 713
2076 TCTTGCGTAAAGAACCATCCATGCCATATATGACAGGGGCAATGACTTAC 2125
713 InHis.....AsnGlnThrTyrIleSerLeuProCysLysLeuArg 726
2126 GACACCTCGGAGAGTCTAGACCTCCCTTAAGATT..... 2160
727 ProCysSerLeuSerTyrValProThrGlnIleProValLeuPheSer 743
2161 .....TTGTCAGAGAAATCTGACTCCCTTGTTCCTTCAATGTC 2201
743 LysnLeuSerTyrThrHisThrAspAsnLeuLysThrLysTyrThrT 760
2202 TCGGTTTGCTTATGGCCATACCGACAATACATGACCAAGATACACTG 2251
760 hTyrProThrValLysGlySerTyrGlyAsnAspSerPheAlaLeuGlu 776
2252 GCTATTCCTCCTGTTAAGGGAAGCTGGGAATGATGCTTCGCTATGAA 2301
777 PheGlyGlyArgAlaProIleCysLeuAsp...GluSerAlaLeuPhe 792
2302 TGTGAGAGAGCTATCCCGTATGCTTTCAGAGCTCCGCTTGTGGTGA 2351
792 uGlnTyrMetProPheMetLysLeuGlnPheValTyrAlaHisGlnGlu 809
2352 TACCCACACAGCCATTTCTTAACCTAGATGATCATGACATAGAAATG 2401
809 LysPheLysGluGlnGlyThrGlnAlaArgGluPheGlySerSerArgLeu 825
2402 ACTTAAAGAAACGCGACAGAGAGCCGCTTCTTCCAAGAGTAAGACTTC 2451
826 ValAsnLeuAlaLeuProIleGlyIleArgPheAspLysGluSerAsp 842
2452 TTCAATCTAGCGGTTCTGTAGGATTAATAATTGAATAATTCCTCGAT 2499
842 sGlnAspAlaThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuVal 859
2500 ....AAGCTACAGATATGATCTCTCCATAGCTTACGTCCTCCGATGATTC 2545
859 rGSerAsnProAspCysThrThrThrLeuArgIleSerGlyAspSerTyr 875
2546 GTAATGATCCAGGCTGCCAGCAACTCTTATGTTCTGTGGGATTTCTTG 2595
876 LysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAla 892
2596 TCGACATGTTGTACAGCTTGTCTAGACAAAGCTCTTGTGACGTGCTGG 2645
892 LysnHisPheCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSer 909
2646 AAATCATCATCTCTTCTCTCAAACTTTGAAGTTTTCAGTCAGTTGTAAG 2695
909 heGluLeuArgGlySerSerArgAsnTyrAsnValAspLeuGlyAlaLys 925
2696 TCGAGTTGCGAGGTTCTTCTGTAAGCTATGCTATGCAATTTGGAGGAGA 2745
926 TyrGlnPhe 928
2746 TTGCGATTT 2754
seq_name: gb_pat:A01851
seq_documentation_block: 3000 bp DNA PAT 21-JAN-2000
LOCUS A01851
DEFINITION Sequence 25 from Patent W09858953.
ACCESSION A01851
VERSION A01851.1 GI:6731876

```


KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3000)
AUTHORS Madsen, A. and Birke Lund, S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 30-DEC-1998;
MADSEN ANNA SOFIE (DK); BIRKE LUND SVEND (DK)
FEATURES
source
1..3000
/organism="unidentified"
/db_xref="taxon:32644"
259..>3000
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69200.1"
/db_xref="GI:673187"
/translation="MKSQSEMLVLSSTLACFTSCSTVAATAENGPSQSPDSTNTG
TTPKNTTGTIDYTLTGDTLTONLSDSALITGCFSDITTESLFGKGYSLFNLKIS
SNEGALVTTIDKNSLTGFSLLTLEAIVSSVITTPSGGAVKCGDILFDNNGILF
KODYCENGGALISTKNSLTKNSTGSISEGNKSATGKKGAIICATGVDITNNAPT
LFSNNIAEAGCAINSTGNCITGTNTSLVSENSYTAAGNGLSCADADYISGNOS
VFESGNOAVANGALYAKKLITLASSGGGGSIFSNINVOGTNGGALISLAAGCSL
SAEAGDITENGNAIVATPOTKRNSIDIGSTAKTINLRISGHSIFPYDPTANTLA
DSTDTLNKAKAGNSIDYSGSIVSGEKLSEDAKAVADNLSTIKOPVTLTACNLTA
KRGVLDTRGFTQTAGSSVIMDAGTTLKASTEEVLTSLISVDSLGEKKVIAASA
ASKNVALSGPILLDNOGNAYENHDLGKTDFSVOLSLGTAITTDVPAETVATPT
HYGDTGWGMWVDIATSPKTKATLAWNTGILPNPERQGLVPAVNSLWSFSDIOA
IOGVIERSLATLCSDRGFAAGVANEFLDKKERKRYRHKGGYALIGAAQTCSEN
ISFAFCOLFSGDKDELVAKNHDTYAGAYIIOHIECSGFIQCLDKLPQSWSHKPLV
LEGOLATSHVSNLDTKRYTAYPEVKGSGMGNMFMMLCASSHSEYELHCPPTVAPYI
KLNTYIRSDSEKTEGSRPDSNLEPLSPICVKEKESDCNDFSDILTSVPP
LIRNDPKCTTALVIGASWETAYANNLAROLDVRAGSHIATSPMEVLGQFVEVGRS
"

BASE COUNT 867 a 597 c 658 g 878 t
ORIGIN

alignment_scores:
Quality: 2000.00 Length: 936
Ratio: 2.801 Gaps: 20
Percent Similarity: 76.282 Percent Identity: 46.261

alignment block:
US-09-428-122-2 x AB1851 ..

Align seg 1/1 to: AB1851 from: 1 to: 3000

1 MetIyserSerPheProLySPhaValPhe...SerThrPheAlaIlePhe 16
|||||.....
259 ATGAATCGCAATTTTCCTGCTAGTCTCTCTCGACATTGGCATTTGTT 308
16 e.....ProIeuSerMetIleAlaThrGluThrValIeuAspSerS 30
|.....
309 TACTAGTTGTTCCACTGTTTTCCTGCACTGCGAATAATAGCCCTT 358
30 eHlaSerPheAspGlyAsnLysasn...GlyAsnPheserValArgGlu 45
|||||.....
359 CTGATACCTTTGACGGAAGTACTACACAGGCACTTACTCCCTAAAT 408
46 SerGlnGluAspAlaGlyThrThrIlePheLysGlySerValThrIle 62
|||||.....
409 ACG.....ACTACTGGAATAGACTATACCTGACAGAGATATACTCT 452
62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
|||||.....
453 GCAAAACCTT...GGGATTCGGCAGCTTTAACGAAAGGTGTTTCTG 499
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuPhe 95
|||||.....
500 ACACTACGGAATCTTTAAGCTTTGCCGCTAAGGGTACTACTCTTTCTTT 549

96 GlnThrValAspAlaGlyThrValAlaIleGlyAlaValAsnSerSerVal 112
|||||.....
550 TTAATAATTAAGTCT...ACTGCTGAAGCGGCAGCACTT...TCTGTAC 593
112 ValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIle 129
|||||.....
594 AACTGATATAAATCTGCTGCTACAGAGATTTTGAGTCTTACTTTCTTAG 643
129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
|||||.....
644 CGGCCCATCATCGATATACACAACCCCTCAGGAAGAAGTGCAGTTAA 693
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
|||||.....
694 TGTGGA...GGGATCTTACATTGTATACAAATGAACTATTTATTTAA 740
160 rLysAsnPheserThrAspAsnGlyAlaIleThrAlaIleThrLeuS 177
|||||.....
741 ACAAGATTACTGTGAGAAATGCGGAGCATTTTACCAGAAATCTTT 790
177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSer 193
|||||.....
791 CTTTGAATAAACACACGCGATCGATTTCTTTGAAGGAATAAATCGAGC 840
194LysLysGlyAlaIleGlnThrSerAspAlaLeuThrI 207
|||||.....
841 GCAACAGCGGAAAGGCTGCTATTGTCTACTGTACTATAGATAT 890
207 eThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSer 224
|||||.....
891 TACAATAATATACGGTCCTACCTCTTCGAAACATATTTGCGAAGCTG 940
224 erGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
|||||.....
941 CAGGTGAGCTATATAATAGCACAGAACTGTACATTCAGGGAATAGC 990
241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 257
|||||.....
991 TCTCTGTATTTCTGAAATATAGTGTGAC.....CGCACCCG 1028
257 rGlyAspMetSerGlyGlyAlaIleCysAlaThrLysThrSerThr 274
|||||.....
1029 AGCA.....AATGAGAGAGCTCTT.....TCTGAGAG 1057
274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
|||||.....
1058 CCGATGTATACATATCTGGAAATCAGAGTGTACTTCTCAGAAACCAA 1107
291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeuAla 307
|||||.....
1108 GCTGTAGCTAATGCGGAGCATTTATGCTAGAACCTTACACTGCTTC 1157
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGly 322
|||||.....
1158 CGGGGGGGGGGGGATCTCTCTTCTTAACAATATAGTCCAGAGTACCA 1207
322 hrAlaProLysGlyAlaIleAlaIleGluAspSerGlyLysLeuSer 338
|||||.....
1208 CTGACAGTATGTGTGAGCATTTCTATCTGCGACCTGAGAGTATAGT 1257
339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrVal 355
|||||.....
1258 CTTTGACGAGAAGCAGGGGACATTACCTTCATGGGAATGCCATTTGTC 1307
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThr 371
|||||.....
1308 AACTAACCCACAACACTCAAAAATTCATGTGACATAGATCTACG 1357
371 lAlaMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleThrPhe 387
|||||.....
1358 CAAAGATCACGAATTTACGTGCAATATCTGGCATATGATCTTTCTTAC 1407
388 AspProIleThrThrGlySerSerThrThrValThrAspAlaLeuLys 404

```

|||||.....:|||||.....:
1408 GATCGATTACTGCTAATACGGCTGCGATTCTACAGATTACTTAAATCT 1457
404 1asnGluThrProAlaAspSerAlaLeuGlnTyThrGlyAsnIleLeuP 421
|||||.....:|||||.....:
1458 CAATAGAGCTGATGCAGGTAATAGACAGATTATAGTGCGATGTTT 1507
421 heThrGlyGluLeuSerGluThrGlnAlaAlaAspSerGlyAsnLeu 437
|||||.....:|||||.....:
1508 TTTCGGTGAAAAGCTCTCTGAAGATGAAGCAAAAGTGCAGACACCTC 1557
438 ThrSerGlyLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLe 454
|||||.....:|||||.....:
1558 ACTTTCACGCTGAAGCAGCCGTGTAACCTTACTACAGAAATTTAGTACT 1607
454 uLyHisGlyValThrLeuGlnThrGlnAlaPheThrGlnGlnAlaAsp 471
|||||.....:|||||.....:
1608 TAAACGGGTGTCACTCGATACAGAAAGCTTACTACAGCCCGGCTT 1657
471 eArGleuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
|||||.....:|||||.....:
1658 CCTGTATTATATGATGCGGGCACAACGTTAAAGCAAGTACAGAGAG 1707
487 SerThrLeuAsnLeuValIleAsnIleSerSerIleAspGlyAlaGly 503
|||||.....:|||||.....:
1708 GTCACTTAAACAGGCTTTCCATTCCTGTAGACTTTTGGCGAGGATA 1757
503 sLyAlaGlyIleGluThrLyAlaThrSerGlyAsnLeuThrLeuSerG 520
|||||.....:|||||.....:
1758 GAAAGTGTATATGCTGCTCTCTGCGACAGATAAATAGTACCCCTTAG 1807
520 LyThrIleThrLeuLeuAspProThrGlyThrPheTyrgLysAsnHis 536
|||||.....:|||||.....:
1808 GTCCGATCTCTTTTGGATACCAAGGAATGTTATGAATAATCACGAC 1857
537 LeuArGAsnProGlnSerTyrrAspIleLeuGluIleuLyAlaSerGly 553
|||||.....:|||||.....:
1858 TTAGGAAATACTCAGACTTTTCATTTGTCAGACTCTCTCTGGGTAC 1907
553 rValThrSerThrAlaValThrProAspProIleMetGlyGluLysPhe 570
|||||.....:|||||.....:
1908 TGCACAACACTACAGATGTTCCAGCGGTCTCTACAGTACGAACTCTA 1957
570 1stTyrgLysTyrgLysGlyThrTrpGlyProIleValThrPhe... 582
|||||.....:|||||.....:
1958 ACATGGGTATCAAGTACTGGGGA...ATGACTTGGGTGATGATACC 2004
583 ...GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrG 598
|||||.....:|||||.....:
2005 GCAAGCACTCCAAAGACTTAAGACAGACATTTAGTGGACCAATACAG 2054
598 yTrpIleProAsnProGluArgIleLeuSerLeuValProAsnSerLeu 615
|||||.....:|||||.....:
2055 CTACCTTCGAAATCTGAGCGTCAAGACCTTTAGTTCCTAATAGCCTT 2104
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyThrLeuMetGluThr 631
|||||.....:|||||.....:
2105 GGGGATCTTTTCAAGACATCCAAAGCATTCAAAGTCTCATAGAGAACT 2154
632 AsnGluGlyLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSer 648
|||||.....:|||||.....:
2155 GCTTTGACTCTTTGTCAGATCAGAGCTCTGGGTCGGGAGTCCGCA 2204
648 nPhePheHisLysAspSerThrLysThrArgArgGlyPheAlaGlnHis 665
|||||.....:|||||.....:
2205 TTTCCTAGATTAAAGATGAAGAAAGGAAAGCAAAATCCCTCATTAAT 2254
665 eArGlyGlyValIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681
|||||.....:|||||.....:
2255 CTGGGAGATGCTATCGAGGTGCAAGCCCAAACTGTCTTCAAAACTTA 2304
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspLysPhe 698
|||||.....:|||||.....:

```

```

2305 ATTAGCTTGGCTTTGGCCACTCTTTGGTAGCGATAAAGATTCTTAGT 2354
698 1AlaLysAsnGlnGlyThrValTyrgLysGlyThrLeuTyrgLysHisA 715
|||||.....:|||||.....:
2355 CGCTAAATAATCATATGATACCATTCATGACAGAGCCTCTATATCCAA 2403
715 snGluThrTyrrIleSerLeuProCysGlyLeuArgProCysSerLeu 731
|||||.....:|||||.....:
2404 .....ATTACGAATGATAGTGGGTTCATAGTGTCTTAGTACT 2442
732 TyrValPro.....ThrIleLeuProValLeuPheSerGlyAsnLe 745
|||||.....:|||||.....:
2443 AAACCTTCGTGGCTCTGGAGTCATAAACCCGTGTTTAGAGGCGAGT 2492
745 uSerTyrrHisHisThrAspAsnAspLeuLysThrLysTyrrThrTy 762
|||||.....:|||||.....:
2493 CGTTTATACCAAGTCAGTAATGATCTGAAGCAAAAGTATACTGCGTAT 2542
762 rOThrValLysGlySerTrpGlyAsnAspSerPheAlaLeuGluPheG 778
|||||.....:|||||.....:
2543 CTGAGTGAAAGGTTCTTGGGGAATATGCTTTTAACATGATGTTGGGA 2592
779 GlyArGAlaProIleCysLeuAspLeuSerAlaLeuPheGluGlnTy 795
|||||.....:|||||.....:
2593 GCTTCTCTCATTTCTTATCCCTGAATACCTGATGTTGTTGATACCTAT 2642
795 tProPheMetLysLeuGlnPheValTyrrAlaHisGlnGluGlyPheL 812
|||||.....:|||||.....:
2643 TCCATATCATCAAACTGAATCTGACTATATAGTCTGAGACAGCTTCTG 2692
812 LuGlnGlyThrGluAlaArgGluPheGlySerSerArgLeuValAsnLe 828
|||||.....:|||||.....:
2693 AGAAGAGTACAGAGAGAGATCTTTGATGACAGACCTTTCATATTA 2742
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAs 845
|||||.....:|||||.....:
2743 TCTTATCCATAGGGGTGAAGTTGAGAAAGTCTCTGATTTATATGACT 2792
845 aThrTyrrAsnLeuThrLeuGlyTyrrThrValAspLeuValArgSer 862
|||||.....:|||||.....:
2793 TTCTATATGATCGATTTATCCCTATGTTCCCTGATCTTATCCGCAAT 2842
862 rOAspCysThrThrThrLeuArgIleSerGlyAspSerTrpLysThrP 878
|||||.....:|||||.....:
2843 CCAATGCACTACAGCACTGTATACACGCGCCTCTGGGAAACTTAT 2892
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHis 895
|||||.....:|||||.....:
2893 GCCAATTACTTAAGCAGACAGAGCCCTTGCAAGTGGTGCAGGCACTCA 2942
895 eCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSerPheGluLeu 912
|||||.....:|||||.....:
2943 GCGCTTCTCTCTATGTTTGAAGTGTGCGGCACTTGTCTTTAATGTC 2992
912 rGgGlySer 914
|||||.....:|||||.....:
2993 GTGGATCC 3000

```

seq_name: gb_baz:CPOMP54

seq_documentation_block: 6030 bp

LOCUS CPOMP54

DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.

ACCESSION AJ001311

VERSION AJ001311.1 GI:3255934

KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane

protein 5

SOURCE Chlamydia pneumoniae.

ORGANISM Chlamydia pneumoniae

REFERENCE 1 (bases 1 to 6030)

AUTHORS Knudsen, K.

TITLE Direct Submission

22-JAN-1999

JOURNAL Submitted (29-AUG-1997) Knudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK

REFERENCE 2 (bases 1 to 6030)

AUTHORS Knudsen, K., Madsen, A. S., Mygind, P., Christiansen, G. and Birkelund, S.

TITLE Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of *Chlamydia pneumoniae*

JOURNAL Infect. Immun. 67 (1), 375-383 (1999)

MEDLINE 99081766

FEATURES

source 1..6030
/organism="Chlamydia pneumoniae"
/isolate="CDC/CWL-029/VR-1310"
/db_xref="taxon:83558"
/clone="B8F3"
/clone="DEX2-18"
/dev_stage="elementary body"
/clone_1lb="pbluescript"
/clone_1lb="pex2"
complement(1..2697)
/gene="omp5"
/product="outer membrane protein 5"
complement(<1..2742)
/gene="omp5"
/note="putative lipoprotein"
/codon_start=1
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CA04671.1"
/db_xref="GI:3255935"
/db_xref="SPTREMBL:086163"
/translation="MKSQFSLVLSSTLACFTSCSTVNAATENIPSPDSFGSTNTG
TTPKTTGIDITLGDITLQNLGDSALTLGCESDTTESSFAKGYLSLFLNKS
SAGALSVTTDKNLSTGFSSLTFLAASVITPSSGAVKCGDLPFNNKILF
KODYCENGAISTKNLSKNTGSSISFEKNSATGKGAICAGTVDIINNTPT
LFSNNIAEAGGAINSTGCTTGNTSLFSESVATNGGALSADADITSGNOS
VFSGNOAVANGAIIYAKRLTLASGGGGSISNNIVCGTTAGNGAIIILAAECESL
SAACDITFNNGAIYATPTPOTTRNSIDIGSTAKITNLAISHSIFPTDITANRA
DSTDLNKKADGNSIDYSGSLFSEGLSEDEAVDNLSTLQVTLTNGNLV
KRGVLTDTGFTQAGSVIMAGTTLKASTEBEVLTLGSLIPVDSIGEEKVIAASA
ASKNALSGPILLDNOGNAYENHDLGKQDFSEVLSLGTATTDVPAVPTVAPT
HYGOGTGMETWDDTASPPTKATLATMNGGYLPNPRGGLVNSLMSFSDIOA
IGOVERSALTLCSDRGFWAGVANFLDCKGKERKRRHSGGVAIGGAOTCSNLA
ISAFQCLFGSKDPLVAKNHDYTGARYIQTHTCSFGICLIDKLGSSKRPV
LGGQALSHVSNDLKRTYATPEVKGSGWNNANMMLGASSHSPYELHCFITYAYI
KINLTYIRQDSFSEKGTGERSPDSSNLFNLSPIGKPEKFSQCNDFSYDLTVYPD
LIRNDPKCTTALVISGASWETVANNLRLQALQVRAASHYAFSPMEVLQGFVEVYGS
"

gene complement(1..2804)
/gene="omp5"
complement(2698..2742)
/gene="omp5"
complement(2747..2752)
/gene="omp5"
/note="putative"
complement(2777..2782)
/gene="omp5"
/note="putative"
complement(2799..2804)
/gene="omp5"
2863..5726
/gene="omp4"
2863..2868
/gene="omp4"
/note="putative"
2896..2902
/gene="omp4"
/note="putative"
2905..2955
/gene="omp4"
2905..2955
/gene="omp4"
2905..2955
/gene="omp4"
2905..2955
/gene="omp4"

RBS

sig_peptide

CDS

/note="putative lipoprotein"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CA04672.1"
/db_xref="GI:3255936"
/db_xref="SPTREMBL:086164"
/translation="MKTSLPVLVSSVLAFSCHLOSLANEELLSPDSENGNIDSGT
TPKTSATYSLTGDFEYEPKGTPLSDSCFQTDNLTFLNGHSLTFEGIDAGTHA
GAASTANKNLTFSGFLSPDSSTVTGOGTSLSAGVNLNTRKLVAVAGNS
TADGAIKASPLFTGTSGDALFSSNSSSTGKATATAGARIANTGVRLSIAS
TSGGAIIDEGTCLISNNKFLYFEGNAATTGATONTAASGPELIIISNNLTIRAS
VAETSGAHAKRLALSSGFTFLRNNSATPRGATISDASELSLSEATGNTT
VNTLTGTSIDTPKRNAINIGSNKFTELRAKRNHTIFEDPITSEGTSSOVKINN
GSAGALNPYOGTILFSGEFLADELKVADNLKSFPTOPVSLSGKLLLOKGVLEST
FSOEAGSLGMDSTGLTSTAGSITITNGIVADVDNVDISLIPIPAEDPNEVEG
LNLIDEGNYSHEMHDQLESLIKTVADAVDNDVDSLIPIPAEDPNEVEG
ONNVMTDITATNTEKATATMTKTGFVSPERKSALVNTLNGVFTDRILOQVLEIG
ATGMHKGQFWYSMTNFLHKTGDNRGRGFTSGGYIGSAHPRKDDLPFARCL
FARDKDCFTAHNSRTYGTGLTFEKRSHLQPNYIRLGRKASEAIEKFPREIPLAL
DVQVFSHDNRMETHTYSLPSEBSNCTIAGSIGDLPVLSNPHLRYTFIPOM
KVENYVSQNSFESSSDGRFSIGRLNLNLSIPVAKFVQGDIGDSTYDLSGFVSD
VYRNPOSTATLVMSPDSWKIRGNWLSRQAFLLRGSNNYVNSNCELFGHYAMELRGS
SRNYVDVGTAKLRF"

mat_peptide 2956..5688
/gene="omp4"
/product="outer membrane protein 4"
stem_loop 5698..5726
/gene="omp4"
/note="putative"

BASE COUNT 1760 a 1341 c 1195 g 1734 t

ORIGIN

alignment_scores:
Quality: 2000.00 Length: 936
Ratio: 2.801 Gaps: 20
Percent Similarity: 76.282 Percent Identity: 46.261

alignment_block:
US-09-428-122-2 x CPOMP54/rev ..

Align seg 1/1 to reverse of: CPOMP54 from: 1 to: 6030

1 MetlyserSerPheProlySphValPhe...SerThrPheAlaIlePhe 16
|||||
2742 ATGAATCGCAATTTCTCGTAGTCTCTCTTCGACATTGGCATGTT 2693
16 e.....ProleuserMetIleAlaThrGluThrValLeuSers 30
|
2692 TACAGTTGTTCCACGTGTTTTCGCACTCGTGAATAATAGCCCT 2643
30 erAlaSerPheaspGlyAsnlyAsn...GlyasnPheSerValArgGlu 45
|||
2642 CTGATACCTTTGACGAGACTACACAGCACGACCTTACTCCAAAT 2593
46 SerGlnGluaspAlaGlyThrThrTyLeuPheGlyAsnValThrIle 62
:::
2592 ACG.....ACTACTGGAATAGACTATACTCTGACAGGAGTATACTCT 2549
62 uGluAsnIleProGlyThrGlyThrAlaIleThrTyLeuSerCysPheAsn 79
:::
2548 GCAAAACCTT...GGGATTCGGCAGCTTACAGAAAGGTTGTTTCTCG 2502
79 snThrLysGlyaspLeuThrPheThrGlyAsnGlyAsnSerLeuPhe 95
:::
2501 ACACCTACGGAATCTTTAAGCTTTGCGCGTAAAGGCTACTCATTCTTT 2452
96 GlnThrValaspAlaGlyThrValAlaGlyAlaAlaValAsnSerVal 112
:::
2451 TTAATATTTAAGTCT...AGTCTGAAGGCGCACACCTT...TCTGTTCAC 2408
112 IValaspLysSerThrThrPheIleGlyPheSerLeuSerPheIleA 129

```

2407 AACTGATAAATCTGCGTACAGGATTTTGCAGTCTTACTTTCTTAG 2258
129 lasePrGlySerSerIleThrThr.....GlySGlyAlaValSer 143
129 lasePrGlySerSerIleThrThr.....GlySGlyAlaValSer 143
2357 CGGCCCATCATCTGTAATCACAACCCCTCAGGAAAGGTGCAGTTAAA 2308
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
2307 TGTGGA...GGGATCTTACATTTGATACCAATGAGACTATTTATTAA 2261
160 rLysAsnPheSerThrAspAsnGlyGlyAlaIleThrAlaLysThrLeu 177
2260 ACAAGATTACTGTGAGAAATGCGGAGCATTTCTACCAAGAATCTTT 2211
177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSer 193
2210 CTTTGAATAACACGACGGATTCCTTTTGAAGGAAATAATCGAGC 2161
194 .....LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrI 207
2160 GCACACGAGGAAAAAGTGGGCTATTTCGTACTGCTACTGATAT 2111
207 ethrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSer 224
2110 TACAATTAATACGCTCCTCCTCCTGCAACAATATGCTGAAGCTG 2061
224 erGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
2060 CAGGTGAGCTATTAATAGCACAGGAACGTGTACAAATTACAGGAAACG 2011
241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThrTh 257
2010 TCTCTGTATTCTTCTGAATAGTGTACA.....GCGACCGC 1973
257 rGlyAspMetSerGlyGlyAlaIleCysAlaTyrLysThrSerThAsp 274
1972 AGGA.....AATGAGAGACTCT.....TCTGAGATG 1944
274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
1943 CCGATGTTACCATATCTGGAAATCAGTGTACTTCTCAGGAACAA 1894
291 SerThrThrAlaGlyAlaIleTyrValLysLysLeuLeuAlaSe 307
1893 GCTGTACCTAATGCGGACCATTTATGCTAAGAAGCTTACACTGCTTC 1844
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyL 322
1843 CGGGGGGGGGGGGTATCTCTTTCTTCAAAATATAGTCCAAGGTACA 1794
322 hrLarProLysGlyGlyAlaIleAlaIleGluAspSerGlyGluLeu 338
1793 CTGCAAGTAATGTGGAGCCATTTCTATCTGACGCTGAGAGTGTAGT 1744
339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 355
1743 CTTTCACACAGACGAGGACATTACCTTCAATGGGAATGCCATTGTGC 1694
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSer 371
1693 AACTACACCAACAACATAAAGAAATCTATGACATAGATCTACTG 1644
371 lalysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheT 387
1643 CAAAGATCACGAATTTACTGCAATATCTGGCATACATCTTTTCTTAC 1594
388 AspProLethrThrGlySerSerThrValThrAspAlaLeuLysVa 404
1593 GATCCGATTACTGTAATACGGCTGGGATTCACAGATACCTTAATCT 1544
404 lAsnGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleLe 421
1543 CATAAAGCTGATGACGATTAATAGTACAGATTATAGTGGTGATTTGTT 1494
421 hethrGlyLysLeuSerGluThrGlyAlaAlaAspSerLysAsnLeu 437
1493 TTTCTGTGTAAGAGCTCTGTAAGATGAAGCAAAAGTTGCAACAACCTC 1444
438 ThrSerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSer 454
1443 ACTTCACTACCTGAGACGAGCTGTAAGTCACTAAGTCAAGAAATTTGTA 1394
454 ulysHisGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspS 471
1393 TAACTGTGTGCTACCTCTGATACGAAGAGCTTTACTCAGACCGGGT 1344
471 erArgLeuGlnMetAspValGlyThrThrLeuGluProAla...AspTh 486
1343 CTCTGTCTTATATGATGCGGGCACACCTTAAAGCAAGTACAGAGAG 1294
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
1293 GTCACTTTAAGAGCTCTTCCATTCCTGTAGACTTTAGCGGAGGTAA 1244
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSer 520
1243 GAAGTTGTAAATTGCTGCTTCGCAGCAAGTAAATGTAGCCCTTAGT 1194
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsnHis 536
1193 GTCCGATCTTCTTCTTGGATTAACCAAGCAATGCTTATCAAAATCACGC 1144
537 LeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAlaSerGly 553
1143 TTAGCAAAACCTCAAGCTTTCATTTGTCAGCTCTGCTGCTGATAC 1094
553 rValThrSerThrAlaValThrProAspProIleMetGlyGluLysPhe 570
1093 TGGACAACTACAGATGTCAGCGGTTCTTACAGTACACTCTACGC 1044
570 lTyrGlyTyrGlnGlyThrTyrPglProIleValTyr..... 582
1043 ACTATGGTATCAAGTACTTGGGA...ATGACTTGGTATGATATAC 997
583 .....GlyThrGlyAlaSerThrThrAlaThrPheAsnTyrThrLysThr 598
996 GCAAGACTCCAAAGACTAAGACAGGACATTAAGCTTGGACCAATACAG 947
598 yTyrIleProAsnProGluArgIleGlySerLeuValProAsnSerLeu 615
946 CTACCTTCGAATCTGAGCGTCMAAGACCTTAGTCTCTAATAGCTTT 897
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAla 631
896 GGGGATCTTTTCAGCATCCAGCATTCMAAGTGTCAAGAGAAAT 847
632 AsnGlyLysLeuGlnLysAspArgAlaPheTyrCysAlaGlyLeuSer 648
846 GCTTGACTCTTTGTACAGATGAGGCTCTGCGCGGAGTGCACA 797
648 nPhePheHisLysAspSerThrLysThrArgArgGlyPheAlaGlnHis 665
796 TTTCTTAGTAATAAGATAAAGAGGGAAGCAAAATCCCTCATTAAT 747
665 erGlyLysTyrValIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681
746 CTGGTGATATGCTATCGAGGAGGCAAGCAAACTTGTTCTGAATACTTA 697
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPhe 698
696 ATTAGCTTGGCTTTGCCAAGCTTTGTTAGCAGATAAAGATTTCTTACT 647
698 lAlaLysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHis 715
646 CGCTAAATAATCATATGATACATGATCAGAGACCTTCTATATCCAAAC 598

```

```

715 snGluThrTyIleSerLeuProCysIleuNrpProCysSerLeuSer 731
597 ..... ATTACAGAAATGTAAGTGGTTCTATAGTGTCTCTTAGAT 559
732 TyrValPro.....ThGluIleProValLeuPheSerGlyAsnLe 745
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
558 AAACTCTCGGCTCTTGAGACATAAACCCCTCGTTTGAAGGAGCGT 509
745 uSerTyhHisThrAspAsnAspLeuSerThyIleTyhThrIleTyh? 762
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
508 CCCTTAATAGCCACGCTCAKTAATGATCTGAAAGACAAAGATACGGTATC 459
762 rGhTyhIleGlySerTyrPolYAsnAspPheAlaLeuGluIlePhe 778
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
458 CTGAGGTGAAGGTTCTTGGGGAGATATGCTTTTAACATGATGTTGGGA 409
779 GlYAlaGAlaProIleCysLeuAspGluSerAlaLeuPheGluIleTyr 795
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
408 GGTTCCTCATCTCTTATCTCGAATACCTGATGTTGTAATACCTATGC 359
795 rProPheMetIysLeuGlnPheValTyhAlaIAsnGlnGluIlePheLys 812
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
358 TCATATACATCAACTGATGATCTGACCTATATACGTACAGACACTTCTCG 309
812 IuGlnGlyThrGluAlaArgGluIlePheGlySerSerArgLeuValAsnLeu 828
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
308 AGAAAGGTACAGAAAGAAATCTTTGATATACAGCAACCTCTTCAATTTA 259
829 AlaLeuPProIleGlyIleArgPheAspIysGluSerAspCysGlnAspAl 845
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
258 TCTTTGCTATAGGGGTGAAGTTTGAAGAAAGTCTCTGATTTGAATAGACT 209
845 aThrTyhAsnLeuThrIleuGlyTyhThrValAspLeuValArgSerAsn 862
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
208 TCTTATATATCTGACTTATCTCTATGTTCTCTGATCTTATCCCAATGATC 159
862 rAspCysTyhThrThrIleuArgGlieSerGlyAspSerTyrPheThrPhe 878
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
158 CCAATGACATACAGCACTTGTATACGCCGACCTCTTGGGAAACTTAT 109
879 GlYhTrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisP 895
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
108 GCCAATAACTTACGACACAGCCCTTGCAGGTGCGTGCAGGACGACGAC 59
895 eCysHesAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluAla 912
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
58 GCCCTTCTCTCCTATGTTGAAGTCTCGGCGAGTGTCTTGAAGATTC 9
912 rGlyIleSer 914
      ::::: :::::
8 GTGATCC 1
seq_name: gb-pat:AB1839
seq_documentation_block: 2787 bp DNA PAT 21-QAN-2000C
LOCUS AB1839
DEFINITION Sequence 13 from Patent WO9858953.
ACCESSION AB1839
VERSION AB1839.1 GI:6131870
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 2787)
AUTHORS Madsen,A. and Birkeland,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHILAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 30-DEC-1998:
MADSEN ANNA-SOFIE (DK); BIRKELUND SVEND (DK)
FEATURES
Source
1..2787
Location,Qualifiers
/db_xref="taxon:32644"

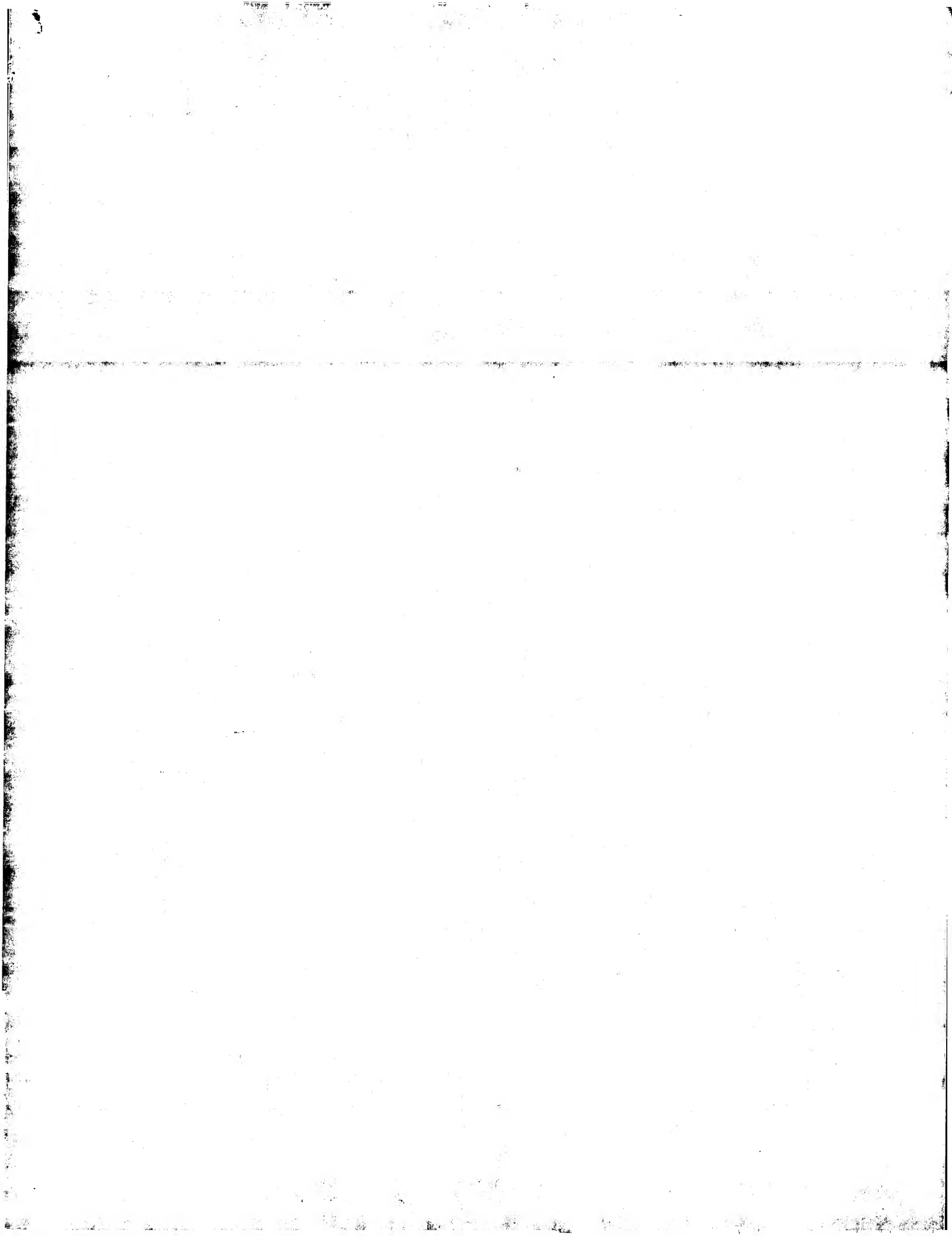
```

BASE COUNT	815 a	689 c	535 g	748 t
ORIGIN				
alignment_scores:				Length: 950
Quality: 1986.00				Gaps: 19
Ratio: 2.770				Percent Identity: 45.263
Percent Similarity: 75.474				
alignment_block:				
US-09-428-122-2 x A81839				
Align seg 1/1 to: A81839 from: 1 to: 2787				
1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr	17			
1 AAGAAATCTCTCTTCATTGGTTGTATGCACTCTTATAGCAGCTTCC	50			
17 OleuSerMet.....IleAlaThrGluThrValLeuA	28			
51 CTTCGACATAAATTCTCTGCGTTTGGCTGCTGTTGTAATCAATCTAG	100			
28 spSerSerAlaSerPheSpLysnLysnAngLysnPhSerValArg	44			
101 GACCTACCAATAGCTCTCTGGA.....CCAGAACTCATACCTCTCA	144			
45 GluSerGngLysAla...GlyThrThrThrLeuPheLysGlyAsnVa	60			
145 GCCCAACACAAACAAATGCAGATGCAGAACTATCTATATCTAACAGGAGT	194			
60 ThrLeuGluAsnLleProGlyThrGlyThrAlaIleThrLysSerCysp	77			
195 CTCATCATCAACCAT...GCAGATCTCCGACAGCTCTAACCGCTTCCGCT	241			
77 heAsnAsnThrLysGlyLysPheLthrPheThrGlyAsnGlyAsnSerLeu	93			
242 TTTAAGAAACACTACGTGGAACTTTCTTCCAGGCCACCGCTACCAATT	291			
94 LeupheGlnThrValAspAlaGlyThrValAlaIleAlaValaLse	110			
292 CTCCTACAAATATGATGACGGGA.....GCGAACTGTCCTTACCA	335			
110 rSerAlValAspLysSerThrThrPheIleGlyPheSerSerLeuSerp	127			
336 TACACTGCAAATTAAGCTTCTCTCTTTCAGGATTCCTATTGTGCAC	385			
127 heIleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSer	143			
386 TAAATCAAAAC.....ACGAATGCTACACAGACAGACAGCATTCAC	429			
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe	160			
430 ..TCCACAGAGAGCTTGTCTATTTCACGCAACATATGTTGCTACTTTGG	476			
160 rLysAsnPhSerThrAspAsnGlyValAlaIleThrAlaLysThrLeuS	177			
477 CCAAAACCTTTCTATGACAAATGGAGGCCCTTCAAGGACGCTATCA	526			
177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer	193			
527 GTCATTCG...CTAAACCCCAACCTAACGTTTGCCAAAACAAAGCAAG	573			
194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAs	210			
574 CAAAAGGGGGGCCCTTATTCACGGAGGAGGATTCAAATTAACATTAAC	623			
210 ngInaLysLysPheSerAspAsnThrSerSerAspSerGlyAlaA	227			
624 GTTAATACACATCATTTCTGAAATATACCGCGGCAACAATGGCGGAG	673			
227 IaIlePheThrGluAlaSerValThrIleSerAsnAlaLysValSer	243			
674 CCATTTCACGGAAGCTAGAGTTTATTCACACACAAAGCAATTAAGC	723			

244	PhellaeaspasnyValThnGlyAlaSerSerSerThrThnGlyaspWe	260
724	TTTAAACCAATAGTGGACCGCAACCTCGACT	756
260	tSerGlyAlaIleCysAlaTyrLysThrSerThrAspThrLysVal	276
757	.ACAGGGGACCATTACGTGATGATGATACATACGCCCAACCACTGT	805
277	..ThrLeuThnGlyAsnGlnMetLeuLeuPheSerAsnAsnSerThr	292
806	TAACTCATCGACACACGGGAACTGAACTTTATAGAAATACAGCAATT	855
293	ThrIleGlyAlaIleTyrValLysLysLeuLeuAlaSerGlyAl	309
856	ACTAGTGTGGGGCGATTATTACTGACAACTAGTCTTCTTCTGGAG	905
309	ylLeuThrLeuPheSerArgAsnSerValAsnGlyThrAlaProLysG	326
906	ACCTACGCTTTTAAAAACAACCTGCTATAGATGACTGCAGCTCCCTTNG	955
326	LysGlyAlaIleAlaIleGlyAspSerGlyLuleuSerLeuSerAlaAsp	342
956	GAGAGCAATTCGCAATTCGCACTCGCACTTGAATCTTTCGGCTCTT	1005
343	SerGlyAspIleValPheLeuGlyAsnThrVal	356
1006	GCTGGACATCACTTTGAAAGAAACACAGTAGTCAAAGAGCTTCTTC	1055
356	rThrProGlyThrAsnArgSerSerIleAspLeuGly..ThrSerAlaL	372
1056	GAGCAGACCACTACCAAGAAATCTATTACATCGGAAACACCAATGCTA	1105
372	ysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheTyrAsp	388
1106	AGATTGACAGCGTGGCAGCCTCCAGCAGCAATACATATCTACTTCTATAT	1155
389	ProIleThrThnGlySerSerThrThrValThrAspValLeuLysValAs	405
1156	CCATATAACAATCAACCATCTGCTAGCTCTCTCAGATGCTCTAACTTAA	1205
405	ngLutThrProAlaAspSerAlaLeuGlnIlyThrGlyAsnIleIlePheT	422
1206	TGCTCTGACCTTGCAGGGAATCTCGCATATACAGAACCATCGATTATT	1255
422	hriGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThr	438
1256	CTGGAGAGAAGCTCTGGGAAGCAGAAAGCTGCAGAGCTGATTACTCAA	1305
439	SerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLeuLy	455
1306	TCATCAATTCAGCAACCTCTAATCTTGGGGAGGGCAACTCTCTTTAA	1355
455	sHisGlyValThrLeuGlnThnGluAlaPheThrGlnIleAlaAspSerA	472
1356	ATCAGAGTCACTCTAGTGGTGAATCTCTTTCCGAATCTCGGGCTCTA	1405
472	rgLeuGlnMetLysValGlyThrThrLeuGluProAlaAspThrSerThr	488
1406	CCCTCTCTATGGATCCAGGGACCACTTATGAAACCGCTGATGGATCACT	1455
489	IleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLysVal	505
1456	ATCAATTAATCTTGTCTCAATCTAATTCCTTAAAGACCAAGAAAGGC	1505
505	AlsIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrI	522
1506	TACGCTAAAGCAACACAGCAAGTCAACAGTCACTTATCTGATGCGC	1555
522	LeThrLeuAsnAspProThrGlyThrThrPheTyrGluAsnHisSerLeuArg	538
1556	TCTCTCTGTGAATCTCTTGGAAATGTCTGGAAGATGTCTCTTGGAAAT	1605

539 AspGdcInsbSerIyrtrpSplLeuGluLeuLysAlaSer.....GlyTh 553
1606 AACCTTAAGTCTTTTGTGTCTCACTCTTACTGTCAGACCCGCCAA 1655
553 rValThSerThAlaValThrProAspProIleMetGlyLysPheH 570
1656 TATTCACATCAGAGACTTATGCTGCTGATCCCTACAAAAAATCTATCC 1705
570 tStyrgLyTYrGInglyThrTPGlyProIleValITrPGlyThelYAla 586
1706 ATGGGGATACCAAGGAATGGGC...TATCTGTGGCAAGAGTACT 1752
587 SerThr.....ThrlathrPheAsnITrPrLysIThnglyTYrII 600
1753 GCACGTAATCCAAAGCAGCAGCTCTTACCTGACAAAACAGGATACAA 1802
600 eProAsnProGluArgIleGlySerLeuValProAsnSerLeuITrPAsnA 617
1803 TCCGAATCCGAGCGCTGCGAAGCTTGTGGTACCAAGCTATGGGGAT 1852
617 lAbhelLeaSplIesSerLeuHstIyrLeuMetGluThAlaSnGlu 633
1853 CCTTGTGTGTGGCTCCATACACAGCTTGATGCCACTAAAGTACGC 1902
634 GlyLeuGInglyAspArgAlaPheITrPCysAlaGlyLeuSerAsnPheP 650
1903 CAATCTCAAGAAACCTGGCGGCACTCGTGATGGAAGGATCTGCAGCTTCT 1952
650 eHISLysAspSerIThrLysThArgArgGlyPheArgHISLeuSerGlyG 667
1953 CCATTAAGATACACGAGATATAAAGGTTTCGGCCACATAAGTACAG 2002
667 lYTrValIIleGlyLysLeuHstIhrCysSerAspLysIleLeuSer 683
2003 GTTATGTTAGAGCAGACTCAACATAGCTTCTGATAATCTATACCT 2052
684 AlAlaIaPheCysGInLeuPheGlylYArgAspArgAspITrPheValAlaTy 700
2053 GCAGCCTTGCCAAATATTCGGGAAATAGACATCACTTTAATAATA 2102
700 sAsnGInlyThrValIYrGlyGlyThLeuTYrGInHISAsnGluIT 717
2103 AATAGAGCTTGCTGCTATGCAGCTCTCTTCATCTCCAGCATCTAGCA 2152
717 hITrYrIleSerLeuProCysLysLeuITrProCysSerLeuSerITrVal 733
2153 CCTTGCTTCT.....CCAAGCTGTAGCGTACCTT 2184
734 Pro.....ThGluIIleProValLeuPheSerGlyLysLeuSerITr 747
2185 CCGTGAATCAAGAGGAGCAGCCGCTCTCTTGATGCTAGATCAGCTA 2234
747 rThrHISIThrAspAsnSplLeuLysIThrLysITrThTYrProThY 764
2235 TATCTATAGTAAATACTATGAAACCTTATACACCACAGCACAAG 2284
764 alLYGlyISerITrPGLysAspSerPheAlaLeuGInPheGlyGlyArg 780
2285 GAGAAAGCTCGGTATATATACCGTGGCCCTGTGAAACTTGGAGCTCC 2334
781 AlaPro.....IleCysLeuAspGluSerAlaLeuPheGluGInIlyrMeITr 796
2335 CTACACACACTGCTTTAAGCCATGAGAGGCTCTTCCACGCGATTTTCC 2384
796 oPheMetLysLeuGInPheValITyAlaHISGInGlyGlyPheLysGluG 813
2385 TTTCACTCAAAGTAAAGACTTCGTACATACCAAGATAGCTTCAAGAAC 2434
813 InGlyIThGlu...AlaArgGInPheLysSerArgLeuValAsnLeu 828
2435 GTATACACTCTGTGATAGATCTTTGATAGCGGTATTTATTTAAGGTC 2484
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGInAspAl 845

```
2485 TCCTGCTTATTGGATTACCTTCGAGAGATTCTCGAGAAACGAGCGCTGC 2534
      :::::::::::::: :::::::::: :::: ||
845 aThrTYrAsnLeuThrLeuGlyTYrThrValAspLeuValArgSerAsnF 862
      :::::::::: :::: :::::::::::::: :::: |||
2535 GTCTTAGAGAGCTACTGTCAATCTACGTTGCGGATGTCTATCGTAGAATG 2584
      |||
862 roAspCysThrThrThrLeuArgIleSerGlyAspSerTrpLysThrPhe 878
      |||
2585 CTGACTGTCAGACAGCTCTCTTAATCAACAATACTCTGTAACAACTACA 2634
      |||
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgIleGlyAsnHisPh 895
      |||
2635 GGAACGAATCTCTCAGACAAAGCTGTATCGAAGAGCAGGGATCTTTTA 2684
      :::::::::::::: :::: ::::::::::
895 eCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSerPheGlnLeuA 912
      :::::::::: :::: :::::::::::::: :::: |||
2685 TGCCTTCTCCAAATCTTGAGGTCAACAGTAACCTATCTATGSAATTC 2734
      |||
912 rGcLYSerSerArgAsnTYrAsnValAspLeuGlyAlaLysTYrGlnPhe 928
      |||
2735 GTGGATCTTCACGACGCTACAAATGACAGATCTTGAGGTAAGTTCCAGTTC 2784
```



Sequence 3000 BP; 871 A; 617 C; 634 G; 878 T; 0 other;

alignment_scores:

Quality: 4774.00 Length: 928
 Ratio: 5.144 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-428-122-2 x A27021 ..

Align seg 1/1 to: A27021 from: 1 to: 3000

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePheBr 17
|||||
101 ATGAAGTCTTCTTCCCAAGTTGTGATTTTCTACATTTGCTATTTTCCC 150
17 OleuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPhe 34
|||||
151 TTGTGCTATGATTGCTACCGAGACAGCTTTGGATTCAAGTGGAGATTGG 200
34 SPGLYAsnLysAsnGlyAsnPheSerValArgGluSerGlnGluAspAla 50
|||||
201 ATGGGAATTAATAATGTAATTTTTCAGTTCCTGAGAGTCAGAGAGATGCT 250
51 GlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAsnIleProG 67
|||||
251 GGAAGTACTACTACTTATTAAGGAAATGTCAGCTTAGAAAATATTCCCTGG 300
67 YThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
|||||
301 AACAGCGACAGCAATCACAAAAGCTGTTTAAACAACACTAAGGCGCAT 350
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
|||||
351 TGACTTTACAGGTACGGGAAGCTCTATTGTTCCAAACGCTGATGCA 400
101 GlyThrValAlaGlyAlaAlaValAsnSerSerValValAspLysSerTh 117
|||||
401 GCGAGCTAGACAGGGGCTGCTGTTTAAACGACCGTGAATTAATCTAC 450
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
|||||
451 CACGTTTATAGGGTTTCTCTGCTATCTTTATTGCGCTCTCGAAGTT 500
134 erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
|||||
501 CGATTAACACCGGCAAGAGACCGTTACGCTCTACGGGAGCTGATAGT 550
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAs 167
|||||
551 TTGACAAAATAATGTCAGTTGCTCTCAGCAAAAACCTTTCAACGAGATA 600
167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
|||||
601 TGGGGGTGCTATCACCGCAAAAGCTCTTTCATTAAACGAGGACTACAATGT 650
184 erAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
|||||
651 CAGCTGCTGTTTCTGAATAATACCTCTCAAGAAAGCGGAGCATTTGAG 700
201 ThrSerSerAlaLeuThrIleThrGlyAsnGlnGlyGluValSerPheS 217
|||||
701 ACTTCGATGCGCTTACCATTAAGTGAACCAAGGCGAAGCTCTCTTTTTC 750
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
|||||
751 TGACAATACTTCTTCGATTTCTGAGAGCTGCATTTTACAGAGGCTCTCG 800
234 aIThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
|||||
801 TGACTATTCTATAATGCTAAAGTTCTCTTTATGACAAATAGGCACAA 850
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaAlaIleCysAl 267
|||||
851 GGAGCGAGCTCTCAACAAAGGGGATATGTCAGGAGGTGCTATCTGTGC 900

```

```

267 aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
|||||
901 TTATAAAAGTGTACAGATACATACTAAGTCTACCTCTACGAAAATCAGATGT 950
284 euleuPheSerAsnAsnThrSerThrAlaGlyAlaIleTyrVal 300
|||||
951 TACTCTTCAGCAACATACATACGACACAGCGGAGAGAGCTATCTATGTG 1000
301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerThrArgS 317
|||||
1001 AAAAGCTCGACAGCTGCTCCGGAGAGACTTACCTTATTCATGTAAGATAG 1050
317 rValAsnGlyGlyThrAlaProLysGlyAlaIleAlaIleGluAspS 334
|||||
1051 TGTCATGAGAGTAGACAGCTCTTAAGGTGAGGACATAGCATTCAGAAATA 1100
334 ergLysGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
1101 GTGGGGAATTGAGTTTATTCGCCGATAGTGGTGCATTTGCTTTTAAAGG 1150
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
|||||
1151 AATAAGTCACTCTCTACTACTCTCGGAGCAAGATAGAGATAGATGCACTT 1200
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
|||||
1201 AGGAGAGAGTGCAAAAGATAGACAGCTTGCGTTCTGCTGCTGATAGCCCA 1250
384 leTyrPheTyrAspProIleThrThrGlySerSerThrThrAlaThrAsp 400
|||||
1251 TCTACTTCATATGATCCCATACTACAGATCATCCAAACATTTTCAGAT 1300
401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
|||||
1301 GTCTTAAAGTTAATGAGACTCCGCGAGATTTCTGCATCAATATATACAGG 1350
417 yAsnIleIlePheThrGlyGlyLysLeuSerLysThrGluAlaAlaAspS 434
|||||
1351 GAACATCATCTTCACAGAGAAAAGTTATCAGACAGAGGCGCGAGATT 1400
434 erLysAsnLeuThrSerLysLeuGlnProValThrLeuSerGlyGly 450
|||||
1401 CTAAAATCTTACTTCGAAGCTACTACAGCCTGTAACTCTTGACGAGAGT 1450
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
|||||
1451 ACTTATCTTTAAACATGAGAGTGACTCTGCAGACTCAGGCATTCACATCA 1500
467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluPro 484
|||||
1501 ACAGCAGATTCCTGCTCGAAATGAGAGTAGGAATCTCTAAGACCTG 1550
484 IaAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
|||||
1551 CTGATACAGCACCAATTAACATTTGGTCATTAACTACATGCTTCTATAGAC 1600
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
|||||
1601 GGTGCAAAAGAGGCAAAAATAGAAACCAAGCTACGTCAAAATAATCTGAC 1650
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
|||||
1651 TTTATCTGAACCAATCACTTATTTGACCCGACGCGGACGTTTATATAAA 1700
534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
|||||
1701 ATCATAGTTTAAGAAATCTCTAGCTCAGACATCTTAAGAGCTCAAAAGCT 1750
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
|||||
1751 TCTGGAAGTGTAAACAGACCGCGAGTGAATCTCAGATCTTAAATGGGTGA 1800

```

```

567 ulyspHeHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGlyTyr 584
|||||
1801 GAAATTCATTACGCTATCAGGAACTGGGGCCCAATGTTGGGGGA 1850
584 hrGlyAlaSerThrThrAlaThrPheAsnTrpTrpLysThrGlyTyrIle 600
|||||
1851 CAGGGGCTTACAGACTGCACACTTCAACTGAGATAAAGCGGCTATATT 1900
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
1901 CCTATCCCGAGCGTATCGGCTTTAGTCCCTATATAGCTTATGGAATGC 1950
617 aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaSnGluG 634
|||||
1951 ATTTATGATATATAGCTCTCTCCATATCTTATGAGACTGCACAACGAG 2000
634 lyleuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAspPhe 650
|||||
2001 GGTGCAAGGAGACCGTCTTTGGTGCTGCTGATTAATCTTCTG 2050
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
|||||
2051 CATAGAGATAGTACAAAACACGACGGGGTTCGCCATTGAGTGGCGG 2100
667 YTYrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSer 684
|||||
2101 TTATGCTATAGGAGAAACCTACATACTGTTCCAGATPACATCTTAGTG 2150
684 lalaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
|||||
2151 CTGCATTTTGCAGCTCTTGAGAGATAGAGACTACTTGTGAGCTAG 2200
701 AsnGlnGlyThrValIYrGlyGlyThrLeuTyrTyrGlnHisAsnGluTh 717
|||||
2201 AATCAAGGTAAGCTACGAGGAACCTCTATPACAGCAACGAAC 2250
717 rTYrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
|||||
2251 CTATATCTCTCTCTTGCACAACTACGCGCTTGTCTGCTTATATGTC 2300
734 roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
|||||
2301 CTACAGAGATTCCTGTTCTCTTTCAGAAACCTTACCTACACCATAAG 2350
751 AspAsnAspLeuLysThrLysTyrThrThrTyrProThrValLysGlySe 767
|||||
2351 GATPACGATCTGAACCAAGATATACACATATCTCTACTGTTAAGGAG 2400
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyIleValArgAlaProIle 784
|||||
2401 CTGGGGAATGATAGTTTCCGCTTAGAATTCGGTGAAGAGCTCCGATT 2450
784 yLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
|||||
2451 GCTTAGATGAAGGCTCTATTGAGCAGTACATGCCCTTCATGAAATG 2500
801 GluPheValIYrAlaHisGlnGlyPheLysGlnGlnGlyThrGluVal 817
|||||
2501 CAGTTTCTATGACATCAGGAAGGTTTAAAGAACAGGGAAGAGAGAGC 2550
817 aArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
|||||
2551 TCGTGAATTTGGAAGTAGCCGCTGTGTGAATCTGCTTACCTTAGCGGA 2600
834 leaArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
|||||
2601 TCCGATTGTGAAGAAATCAGACGCAAGGTCACACGTAACCTACACT 2650
851 leuGlyTyrThrValAspLeuValArgSerAsnProAspCysTrpTrpThr 867
|||||
2651 CTGCGTATATACGTGAGATCTTGTCTGAGTACAGCCGACGCTGACGAAAC 2700
867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884

```

```

|||||
2701 ACTGCGAATTAAGCGGTATCTTGAAACCTTCGGTACCAATTTGGCAA 2750
884 rGlnAlaLeuValIleuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
|||||
2751 GACAGCTTTAGTCTTCGTCGAGGAGACCATTTTGTCTTAATCTCAAT 2800
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgas 917
|||||
2801 TTGGAAGCTTTAGCCCAATTTCTTTGAATTCGTCGGGTCAATCGCAA 2850
917 nTYrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
|||||
2851 TTAACAATGTAGACTTGAAGACCAAAATPACCAATTC 2884

seq_name: /SID56/gcdata/geneseq/geneseqn/NA1999.DAT:X06820
seq_documentation block:
ID X06820 standard; DNA; 2787 BP.
XX
AC X06820:
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp8 DNA.
XX
KW Omp8: outer membrane protein 8; surface exposed protein; antigen;
KM infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-R2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB: W88421.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 52-53; 115pp; English.
XX
XX
CC This DNA sequence codes for the novel 90.0 kDa surface exposed
CC protein Omp8 (see W88421) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp15 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15
CC the other, and encode polypeptides of about 89,6-100,3 kDa and
CC about 56,1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA

```

CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.
 XX

50 Sequence 2787 BP; 811 A; 583 C; 598 G; 795 T; 0 other;

alignment_scores:

Quality: 4760.00 Length: 928
 Ratio: 5.135 Gaps: 0
 Percent Similarity: 99.892 Percent Identity: 99.677

alignment_block:

US-09-428-122-2 x X06820 ..

Align seg 1/1 to: X06820 from: 1 to: 2787

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePheR 17
  |||
1 ATGAACTCTCTTCCCAAGTTGTATTCTTCAATTGCTATTTTCCC 50
17 oleuSerMetIleAlaThrGluThrValIleuAspSerSerAlaSerPhe 34
  |||
51 TTGTGTATGATTGCTACCGACAGACTTTTGATTCAGATGCGAGTTG 100
34 spcLysnLysAsnGlyAsnPheSerValArgLysSerGlnLysAspAla 50
  |||
101 ATGGGAATAAAATGGTAATTTTTCAGTTGCTGAGATCGAGAAATGCT 150
51 GlyThrThrTyrIleuPheLysGlyAsnValThrLeuGlnSerProG 67
  |||
151 GGAGACTACCTACTATTAAAGGAAATGTCACCTAGAAAATATTCCTG 200
67 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAsp 84
  |||
201 AACAGGCACAGCATCACAAAAGCTGTTTAAACAACACTAAGGGCGATT 250
84 euThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
  |||
251 TGACTTTCACAGGTACGGGAACCTCTATTGTTCCAAACGGTGATGCA 300
101 GlyThrValAlaGlyAlaValAlaAsnSerSerValValAspLysSer 117
  |||
301 GGGAGCTGTAGCGGGCTGCTGTTACACAGCGGTGTAGATTAATCTAC 350
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySer 134
  |||
351 CACGTTTATAGGGTTTCTTCGCTATCTTTATTCGCTCTCCGGAAGTT 400
134 erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
  |||
401 CCATACACTACCGCAAAAGAGCGGTAGCTGCTACGGGTAGCTTGAA 450
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 167
  |||
451 TTGGACAAAATGTCAAGTTGCTCTTACGCAAAAACCTTTTCAACGGATA 500
167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMet 184
  |||
501 TGGCGGTGTATCACCGCAAAAACCTTTTCAACAGGACTACAAATGT 550
184 erAlaLeuPheSerGlnAsnThrSerSerLysGlyGlyAlaIleGln 200
  |||
551 CAGCTCTGTTTCTGAAAATACCTCTCAAGAAAGCGAGCATTCAG 600
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPhe 217
  |||
601 ACTTCGATGACCCCTTACCTACTGGAACCAAGGGAGGCTCTTTTTC 650
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGlnLys 234
  |||
651 TGACATATCTTCTCGGATTCTGGAGCTGCATTTTTCAGAAAGCTCGG 700

```

```

234 alThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
  |||
701 TGACATATTTCTAATAATGCTAAAGTTTCTCTTATTGACATTAAGCTACA 750
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCysAl 267
  |||
751 GGAGCGAGCTCTCAACAACGGGGGATATGTCAGAGGTCTATCTGTGC 800
267 aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
  |||
801 TTATAAATCTAGTACAGATCTAAGCTACCTCACTGGAATTCAGATGT 850
284 euLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTyrVal 300
  |||
851 TACTCTTCAGCAACAATATACATGACACAGCGGGAGGACTATCTATGTG 900
301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsn 317
  |||
901 AAAAAGCTTCGAAGCTGCTCCGAGAGACTTACCTATTCAGTAAATAG 950
317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGlnAsp 334
  |||
951 TGTCAATGAGGTACAGCTCTTAAAGGTGAGCCATACGATTCGAAGATA 1000
334 erGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
  |||
1001 GTGGGGAATGTAGTTATCCCGCGATAGTGATGATGTCTTTTAA 1050
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAsp 367
  |||
1051 AATACAGTACTCTTACTACTCTCTGGAAGATAGAAATAGATATGACTT 1100
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIle 384
  |||
1101 AGGAACGAGTCAAGAGATGACAGCTTGCGTTCTGCTGCTGTTAAGCCA 1150
384 IeTyrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
  |||
1151 TCTACTTCTATGATCCCACTACAGATCTTCCACAACAGTTACAGATT 1200
401 ValLeuLysValAsnGluThrProLysAspSerAlaLeuGlnTyrThrG 417
  |||
1201 GTCTTAAAGTTAATAGAGCTCCGGCAGATTCTGCATCAATATACAG 1250
417 yAsnIleIlePheThrGlyGlyLysLeuSerGluThrGlnAlaIleAsp 434
  |||
1251 GAACATCATCTTACAGAGAGAAAGTTATCAGAGACAGAGCCGACAGATT 1300
434 erLysAsnLeuThrSerLysLeuGlnProValThrLeuSerGlyGly 450
  |||
1301 CTAAAAATCTTACTTCGAAGCTACTACAGCCGTGACTCTTACAGAGAGT 1350
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
  |||
1351 ACTCTATCTTTAAACATGAGGTGACTCTGCAGACTCAGGATTCACATCA 1400
467 nGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlnPro 484
  |||
1401 ACAGGCAATCTCTGCTCGAAATGAGACTAGGAACCTCTATGAACTGT 1450
484 IAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
  |||
1451 CTGATACTAGCACCATTAACAATTTGGTCATTAACATCAGTTCTATAGAC 1500
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeu 517
  |||
1501 GGTGCAAGAGGCAAAATAGAAACCAAGCTACGTAATAAAATTCGAC 1550
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGln 534
  |||
1551 TTTATCTGGAACCATCACTTTATTGAGCCGACGCGACGTTTATGAAA 1600
534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550

```

```

|||||
1601 ATCATAGTTTAAAGAAATCCTCAGTCTACGACATCTTAGAGCTCAACCT 1650
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMeGlyG1 567
1651 TCTGGAACTGTACAGACACCGCACTGACTGCAGATCCTATATATGGGGA 1700
567 ULysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGlyT 584
1701 GAAATTCATTACGGCTATCAGGGAACCTGGGGCCCAATTGTTGGGGA 1750
584 hnglyAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrIle 600
1751 CAGGGGCTTTTACGACTCAACTCAACTGGACTTAAACCTGGCTATAT 1800
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
1801 CCTATCCCGACGCTATCGGCTCTTAGTCCCTATACCTTATGGAATGC 1850
617 aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGluG 634
1851 ATTATATAGATATATACCTCTCCATTAATCTATGAGAGCTGCAACGAG 1900
634 1LysGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
1901 GGTTCAGAGGACCGCTCTTTTGTGTGCTGATTAATCTACTTCTTC 1950
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG1 667
1951 CATAGAGTAGTACAAACACGAGCGGGTTCCCATTTAGAGCGCGG 2000
667 YThrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
2001 TTATGTCATAGAGGAAACCTACATCTGTTCAGATTAAGATTCTTAGG 2050
684 1aAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
2051 CTGCACTTTTGCAGCTCTTGGAAAGATAGAGACTACTTGTACTAG 2100
701 AsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGluTh 717
2101 AATCAAGGTACAGTCTACGAGGAACTCTATATACCAACGAAACGAAAC 2150
717 rTyrlleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
2151 CTATATCTCTCTCTCTGCAAACTAGCGCTTGCTGCTTATATGTC 2200
734 rOthrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
2201 CTACAGAGATTCTCTCTCTTTCAGGAAACCTTAGCTACACCATACG 2250
751 AspAsnAspLeuLysThrLysTyrThrTyrProThrValLysGlySe 767
2251 GATAACGATCTGAAACCAAGATATACACATATCTACTGTTAAAGAGAG 2300
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyGlyArgAlaProIleC 784
2301 CTGGGGGATGATGATTCGCTTAGAATTGCGTGAAGAGCTCCGATTT 2350
784 yLysLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
2351 GCTTAGATGAAGAGCTCTATTTAGAGAGTACATCCCTCATGAAATVG 2400
801 GlnPheValTyrAlaHisGlnGlnGlyPheLysGlnGlnGlyThrGluAl 817
2401 CAGTTGTTCTATAGCACATCAGAGAGGTTTAAAGAACGAGGAAACGAG 2450
817 aArgGlnPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
2451 TCGTGAATTTGGAATACCGCTTGTGATCTTCTTACTATACGCGGA 2500
834 1eArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
|||||

```

```

2501 TCCGATTTGATTAAGGAATCAGACTGCCAAGATGCCAAGTACATCTACT 2550
851 LeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
2551 CTGTTTACTGTGATCTTGTCTGTAAGTACCCGCACTGTACGACAC 2600
867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884
2601 ACTGCAATTTAGCGGTGATTTCTTGAAACCTTCGTACGAAATTTGGCA 2650
884 rGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
2651 GACAGCTTTTATGCTTCCTGTCAGAGAACCAATTTTGCTTTAACTCAAT 2700
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgIleSerSerArgAs 917
2701 TTGAGCGCTTTTACCAATTTCTTTGAAATGGCGTGATCTCCGAA 2750
917 nTyraSnValAspLeuGlyAlaLysTyrGlnPhe 928
2751 TTACAATGTAGACTTAGAGCAAAATACCAATTC 2784
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X91990
seq_documentation block:
ID X91990 standard: DNA: 1230025 BP.
XX
AC X91990;
XX
DT 13-SEP-1999 (first entry)
XX
DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998: 98MO-IB01890.
XX
PR 04-NOV-1998: 98US-0107078.
PR 21-NOV-1997: 97FR-0014673.
XX
PA (GENSET) GENSET.
XX
PI Griffiths R:
XX
DR MPI: 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Claim 1, Page 291-611. 1912pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC pneumoniae, and encodes proteins Y34584-Y35879. C. pneumoniae causes
CC respiratory disease such as pneumonia and bronchitis and is thought
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC encoded by the open reading frames of the C. pneumoniae genome (see
CC Y34584-Y35879) can be used in immunogenic compositions as vaccines.
CC Vectors containing C. pneumoniae nucleotide sequences can also be
CC used as immunogenic compositions, especially where the vector directs
CC the expression of a neutralising epitope of C. pneumoniae.
XX
SO Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
alignment_scores:

```

Quality: 4733.00 Length: 929
 Ratio: 5.100 Gaps: 1
 Percent Similarity: 99.892 Percent Identity: 99.892
 Alignment_block:
 us-09-428-122-2 x X91990 ..
 Align seg 1/1 to: X91990 from: 1 to: 1230025

```

1 MetLysSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
  |||||||
31758 ATCAAGTCTTCTTCCCAAGTTGATTTTCTACATTGCTATTTCCTCC 31807
17 OleuSerMetIleAlaThrGluThrValIleuAspSerSerAlaSerPhea 34
  |||||||
31808 TTGTGCTATGATGTCTACCGAGACAGTTTGGATTTCAGATGGCAGATTGG 31857
34 spLysAsnLysAsnGlyAsnPheserValArgGluSerGlnGluAspAla 50
  |||||||
31858 ATGGGAATGATAAATGGTAATTTTTCAGTTCTGAGAGTCAGAGATGCT 31907
51 GlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAsnIleProG 67
  |||||||
31908 GGAAGTACCTACCTATTTAAGGAAATGTCACTCTAGAAAATATTCCTGG 31957
67 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
  |||||||
31958 AACAGGCACAGCAATCACAAAAGCTGTTTAAACAACACTAAGGCGAAT 32007
84 euThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
  |||||||
32008 TGACTTTCACAGTAAAGGGAATCTCTATTGTGTCACAAAGGATGATGCA 32057
101 GlyThrValAlaGlyAlaValAsnSerSerValValAspLysSerThr 117
  |||||||
32058 GGGACTGTAGCAGGGGCTGCTGTAAACAGACGCTGATAGATAATCTAC 32107
117 ThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
  |||||||
32108 CACGTTTATAGGTTTCTTCGCTATCTTTATGCGTCTCTCGTAAATG 32157
134 erIleThrThrGlyLysGlyAlaValAserCysSerThrGlySerLeuSer 150
  |||||||
32158 CGATTAAGTACGGGCAAGGAGCCGTACTGCTCTACGAGGATGCTTGAGT 32207
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheserThrAspAs 167
  |||||||
32208 TTGAC.AAAAAATGTCAGTTGCTCTTCAGCAAAAACCTTTCAACGGATAA 32256
167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
  |||||||
32257 TGGGGTGTCTACCCGCAAAACTCTTCTTACACAGGACCTACAAATGT 32306
184 erAlaLeuPheSerGluAsnThrSerSerLysLysGlyAlaIleGln 200
  |||||||
32307 CAGCTGCTTTTCTGAAAATCTCTCAAGAAAGGCGGAGCCATTGACG 32356
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSe 217
  |||||||
32357 ACTTCGATGCCCCCTTACCAATTACTGGAACCAAGGGAAGCTCTTTTTC 32406
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluLasev 234
  |||||||
32407 TGACAATACTCTTCGGATTCTGAGCTGCATTTTACAGAAACCTCCGG 32456
234 alThrIleSerAsnAlaLysValSerPheIleAspAsnLysValThr 250
  |||||||
32457 TGACTATTTCATAATAGCTAAAGTTTCTTTATTGACAAATAAGGTACACA 32506
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysal 267
  |||||||
32507 GGAGCGAGCTCTCAACAAGGGGATATGTCAAGAGGTCTATCTGTGC 32556

```

```

267 aThrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
  |||||||
32557 TTATAAACTAGTACAGATACTAAGGTACACCTCAGCTGAATATAGATGT 32606
284 euLeuPheSerAsnAsnThrSerThrAlaGlyAlaIleTyrVal 300
  |||||||
32607 TACTTTCAGCAACAATATACATGCAACAGCGGAGAGCTATCTATGTG 32656
301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
  |||||||
32657 AAAAAGCTCGAAGCGGCTCCGGAGGACTTACCTATTCCGTAGAAATAG 32706
317 rValAsnGlyGlyThrAlaProLysGlyAlaIleAlaIleGluAspS 334
  |||||||
32707 TGCTAATGAGAGTACACCTCTAAAGCTGAGCATAGCTATGCAAGATA 32756
334 erGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
  |||||||
32757 GTGGGGAATTAGTTTATCCGCCGATAGTGTCATGTCTTTTAAAGG 32806
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerIleAspLe 367
  |||||||
32807 AATACAGTCACTTCTACTACTCTCGGGACGAATAGAAAGTATGTCAGCT 32856
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIleArgAla 384
  |||||||
32857 AGGACAGAGTGCARAGATGACAGCTTGTGCTGTGCTGATGAGCCA 32906
384 leryPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
  |||||||
32907 TCTACTCTATGATCCCATTAAGTACAGATCATCCAAACAGTACAGAT 32956
401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
  |||||||
32957 GTCTTAAAGTTATATGAGATCCGGCAGATTCTGCACTCAATATACAG 33006
417 yAsnIleIlePheThrGlyGluLysLeuSerGluThrGluAlaAlaAspS 434
  |||||||
33007 GAACATCATCTTCACAGAGAAAGTTATCAGAGACAGAGGCCGAGAT 33056
434 erLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
  |||||||
33057 CTAATAATCTTACTTGAAGCTACTACACCCGTGAACCTTTTCAGAGGT 33106
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
  |||||||
33107 ACTCTATCTTTAAACATGAGATGAGTCTGCAGACTCAGAGCTTACTCTA 33156
467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
  |||||||
33157 ACAGCGAGATTCCTCGTCTCGAAATGAGAGTACGAACTACTAGAACTG 33206
484 lAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
  |||||||
33207 CTGATATCTGCAACCATTAACAAATTTGGTCATTAAACATAGTTCTATAAC 33256
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
  |||||||
33257 GGTGCAGAAAGCAAAATAAGAAACCAAGCTAGCTCAAAAAATCTGCAC 33306
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
  |||||||
33307 TTTATCTGAAACCATCACTTTATTGGACCCGACGCGCTTTATGAAA 33356
534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
  |||||||
33357 ATCATAGTTTAAGAAATCTCAGTCTCAGACATCTTAGACCTCAAAAGCT 33406
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
  |||||||
33407 TCTGGAAGTGTAAACAAGCAGCGAGTGTACTCAGATCTCTAATATGGGTGA 33456
567 uLysPheHisTyrGlyTyrGlnGlyThrThrProGlyProIleValThrPol 584

```

```

33457 GAAATTCATTACGGCTATCAGGAACTTGGGCCCATTTGTTGGGGA 33506
584 hrc1yAlaSerThrThrAlaThrPheasnTrpHlyThrGlyTyrIle 600
33507 CAGGGCTTTACGACGCAACTTCACTGACATAAACTGGCTATATT 33556
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
33557 CCTAAATCCACGAGGTATCGGCTTTAGTCCCTAATAGCTTATGAAATG 33606
617 aPheIleAspIleSerSerLeuHISThrLeuMetGluThrAlaAsnGluG 634
33607 ATTATATGATATTAAGTCTCTCCATTAATCTTATGAGACGTGCAACGAA 33656
634 lyeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhphe 650
33657 GGTTCAGGAGACCGTGGTTTGGTGGCTGATTAATCTTACTTCTTC 33706
651 HIsLysAspSerThrLysThrArgArgGlyPheArgHISLeuSerGlyG 667
33707 CATTAAGGATGATACAAACACACGCGGGTTTCGCCATTTGAGTGGCG 33756
667 YTYrValIleGlyGlyAsnLeuHISThrCysSerAspLysIleLeuSerA 684
33757 TTATGTCATAGGAGGAACCTACATACTTGTTCAGATAAATCTTATG 33806
684 lAlaIlePheCysGlnLeuPheGlyArgAspArgAspTYrPheValAlaLys 700
33807 CTCATATTTGTGACGCTCTTGGAAAGATAGAGACTACTTTGAGCTAAG 33856
701 AsnGln GlyThrValTYrGlyTYrHleuTYrTYrGlnHISAsnGluT 717
33857 AATCAAGGATGACTACGAGAACTCTCTATTACACACACGAA 33906
717 hTYrIleSerLeuProCysLysLeuArgProCysSerLeuSerTYrVal 733
33907 CCTATATCTCTCTCTCTGCAAACTACGGCCTTGTCTGTATGCTT 33956
734 ProThrGluIleProValLeuPheSerGlyAsnLeuSerTYrThrHIS 750
33957 CCTACAGAGATTCCTGTTCTTTTCAGAAACCTTAGTACACCCATAC 34006
750 lAspAsnAspLeuLysThrLysTYrThrTYrProThrValLysGlyS 767
34007 GGAATAAGATCTGAAAAACCAAGTATACACATATCTTACTGTTAAAGAA 34056
767 eTYrPGLysAsnAspSerPheAlaLeuGlnPheGlyATrGAlaProIle 783
34057 GCTGGGGGATGATAGTTTCGCTTAGAATTCGGTGAACAGCTCCGATT 34106
784 CysLeuAspGluSerAlaLeuPheGluGlnTYrMetProPheMetLys 800
34107 TCGTTGATGTAAGGTGCTCTATTGAGCACTACATGCCCTTCATGAATT 34156
800 uGlnPheValTYrAlaHISGlnGluGlyPheLysGlnGlyTYrHleuAla 817
34157 GCAGTTTGTATGACATCAGGAAGTTTAAAGAACAGGACAGAG 34206
817 lAaArgIlePheGlySerSerArgLeuValAsnLeuAlaLeuProIleGly 833
34207 CTCGTGAATTGGAGAGTAGCGCTTGTGAATCTTGCCCTTACTATGGG 34256
834 lIleArgPheAspLysGluSerAspCysGlnAspAlaThrTYrAsnLeu 850
34257 ATCCGATTGTGATAGGAATCAGACTGCAAGATGCAACGTACATCTTAC 34306
850 rIleuGlyTYrThrValAspLeuValArgSerAsnProAspCysThrTYr 867
34307 TCTTGGTATACGTGATCTTGTCTGTAGTAACCCGAGATGACGACAA 34356
867 hLeuValGlyLeuSerGlyAspSerTYrLysThrPheGlyThrAsnLeuAla 883

```

```

34357 CACTGCCAATTACGGTGATTTCTTGAAAACTTCCGCTACGAATTTGGCA 34406
884 ArgGlnAlaLeuValLeuArgAlaGlyAsnHISPheCysPheAsnSerAs 900
34407 AGACAGCTTTAGTCTCTGTCGAGGAAACATTTTGTCTTATACACAAA 34456
900 nPheGluAlaPheSerGlnPheSerPheGluLeuArgIleSerSerArgA 917
34457 TTTTGAAGCCTTATAGCCAAATTTCTTGAATTCCTGGGTCATCTCGCA 34506
917 sOTyrAsnValAspLeuGlyAlaLysTYrGlnPhe 928
34507 ATTCAATGTAGACTTATGAGGCAAAATACCAATTC 34541

seq_name: /SID56/gcdata/geneseq/geneseqn/NA1999.DAT:X06817
seq_documentation_block:
ID X06817 standard; DNA; 2815 BP.
XX
AC X06817;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.
XX
KM Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
XX
OS Chlamydia pneumoniae.
XX
FH
FT Key location/Qualifiers
FT CDS 1..2787
FT /*tag= a
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX (CHR/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mydind P;
XX
DR WPI: 1999-105610/09.
XX P-PSDB: W88418.
XX
PT Species-specific test for identifying mammals infected with
PI Chlamydia pneumoniae - comprises detecting antibodies specific for
PI outer membrane proteins of C. pneumoniae or nucleic acids encoding
PI these proteins
XX
PS Claim 6; Page 42-43; 115pp; English.
XX
CC This DNA sequence codes for the novel 97.2 kDa surface exposed
CC protein Omp5 (see W88418) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp5 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used

```


CC in the diagnosis of *C. pneumoniae* infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting *in vivo* expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with *C. pneumoniae*.

XX Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

alignment_scores:

Quality: 2058.00 Length: 950
 Ratio: 2.831 Gaps: 20
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:

US-09-428-122-2 x X06817

Align seg 1/1 to: X06817 from: 1 to: 2815

```

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePh 16
  |||||
1 ATGAATCGCAATTTCTCTGTTAGTCTCTCTCGACATTGGCATGTT 50
  |||||
16 e.....ProLeuSerMetIleAthrGluThrValLeuAspSers 30
  |||||
51 TACTAGTGTTCACACTGTTTCTCTGCACTGCTGAATAATATAGGCCCT 100
  |||||
30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgGlu 45
  |||||
101 CTGATAGCTTTGACGGAGTACTACACAGCAGCATTACTCTTAATAAT 150
  |||||
46 SerGluGluAspAlaGlyThrThrThrLeuPheLysGlyAsnValThrLe 62
  |||||
151 ACG.....ACTACTGGAATAGACTATACTGTGACAGAGATATACTCT 194
  |||||
62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnA 79
  |||||
195 GCAAAACCTT...GGGGAATCGGACAGCTTTAACGAAGGGTGTGTTTCG 241
  |||||
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
  |||||
242 ACACACGGAATCTTTAGCTTTGCCGTAAGGGTACTCATTCTTTCTTT 291
  |||||
96 GluThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSera 112
  |||||
292 TTAATATTAAAGTCT...AGTCTGAAGCGCAGCAGCTT...TCTGTATC 335
  |||||
112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
  |||||
336 AACTGATAAATACTGTCTACACAGATTTTCGAGTCTTACTTTCTTTG 385
  |||||
129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
  |||||
386 CGGCCCATCATCGGTAATCACACACCCCTCAGAAAAGGTGCAGTTAA 435
  |||||
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
  |||||
436 TGTGGA...GGGATCTTACATTGATTAACAATGAACTATTATTAA 482
  |||||
160 lLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeus 177
  |||||
483 ACAAGATTACTGTAGGAAATGGCGGAGCCATTCTACCAAGAACTTTT 532
  |||||
177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
  |||||
533 CTTTGAACCAACGACGCGGATCTTCTTTGAAGGAATAAATCGAGC 582
  |||||
194 .....LysLysGlyAlaIleGluThrSerAspAlaLeuThrIl 207
  |||||
583 GCAACAGGGAATAAGGTGGGCTATTGTCTACTGTACTGTAGATAT 632
  |||||
207 eThrGlyAsnGluGlyValSerPheSerAspAsnThrSerSerAsps 224
  |||||

```

```

633 TACAAATTAATACGGCTCCTACCTCTTCGAAACAATATGCTGAAGCTG 682
  |||||
224 eGlyAlaAlaIlePheThrGluAlaSerValThrLysSerAsnAsnAla 240
  |||||
683 CAGGTGAGCTATTAATATACACAGGAAACTGACAAATTTACAGGAATACG 732
  |||||
241 LysValSerPheIleAspAsnLysValThrGlyAlaIleSerSerThrTh 257
  |||||
733 TCTCTGTATTCTTCTGAAAATAGTGTGACA.....GCCAGCCG 770
  |||||
257 rGlyAspMetSerGlyGlyAlaIleCysAlaLyrThrLysThrSerThrAsp 274
  |||||
771 AGGA.....AATGGAGAGCTCTT.....TCTGGAGATG 799
  |||||
274 hrLysValThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnAsnThr 290
  |||||
800 CCGATGTTACCATATCTGGGAATCAGAGTGAATCTTCTCAGGAACCA 849
  |||||
291 SerThrThrAlaGlyAlaIleLysValLysLysLeuGluLeuAlaSe 307
  |||||
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLys 322
  |||||
900 CGGGGGGGGGGGGTATCTCTCTTTCTTAACAATATATGCTCAAGTACCA 949
  |||||
322 hrAlaProLysGlyAlaIleAlaIleGluAspSerGlyLysLeuSer 338
  |||||
950 CTGCGAGTAATGCTGAGCACCATTCTATATCTGCGCAGAGTGAATG 999
  |||||
339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
  |||||
1000 CTTTACGAGAGAGCAGGGGACATTACCTTCATAGGAAATGCAATGTTGC 1049
  |||||
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSera 371
  |||||
1050 AACTACACCAACAATCTACAAAGAAATTTATGTACATAGAGATCTACTG 1099
  |||||
371 lAlysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleThrPheThr 387
  |||||
1100 CAAGATACGCAATTTAGCGCAATATCTGGCATATACCACTTTTCTTAC 1149
  |||||
388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
  |||||
1150 GATCGAATTACTGCTAATACGCGTCCGGAATTTACAGATCTTAAATCT 1199
  |||||
404 lAsnGluThrProAlaAspSerAlaLeuGluThrThrGlyAsnIleLeP 421
  |||||
1200 CAATAGGCTGATGTCAGGTAATAGTACAGATTATAGGCGTCAATGTT 1249
  |||||
421 heThrGlyGluLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeu 437
  |||||
1250 TTTTGTGTGAAAAGCTCTCGAAGATGACCAAAAGTTGACGACCAACCTC 1299
  |||||
438 ThrSerLysLeuLeuGluProValThrLeuSerGlyGlyThrLeuSerie 454
  |||||
1300 ACTTCTACGCTGAAGCAGCTGTAACTTAACGCAAGAAATTTACTACT 1349
  |||||
454 uLysHisGlyValThrLeuGluThrGluAlaPheThrGluGluAlaAspS 471
  |||||
1350 TAAACGGTGTGCTACTCTGATGCAAGGCTTTACTCAGACCGCGGCTT 1399
  |||||
471 eArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
  |||||
1400 CCTCTGTATTATGATGATCGGCGACACACTTAACCAAGTACACAGGAG 1449
  |||||
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 503
  |||||
1450 GTCACCTTAAACAGGTCTTCCATTCTGTAGACTTTTGGGAGAGGTAA 1499
  |||||
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
  |||||

```



```

472 rgleuGluMetAspValAlaIYThrLeuGlu...ProAlaAspThrSer 487
      ::::: ||||| ||||| ::::: |||||
1494 AAGTCGTAATGAGGAGGACTACTTTGAGCAAGCGCTGAGGGGCTC 1543
488 ThrIleAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1544 ACTCTCAATGGCTTAGCCATTATATAGATTCTTATGATGGACAATAA 1593
504 SALaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerGly 521
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1594 AGCTATCATTTAAGCGGACGCGCAAGTAAGAGATTGGCTTATCAGGGC 1643
521 hrIleThrLeuAsnProThrGlyThrPheTyGluAsnHisSerLeu 537
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1644 CTATCATCTGTAGATGCTCAGGGGAACATATATAGACATCATATCTC 1693
538 ArgAsnProIleSerThrAspIleLeuGluLeuLysAlaSerGlyThVa 554
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1694 AGTCACACAGCAGCTTCTTCCTTATATAGCTTCTGCACAAGAGACAT 1743
554 IThrSerThrAlaValThrProAspProIleMetGlyLysPheHisT 571
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1744 GACTCTACAGATATCCCGATACCCCAATCTAATATCTAGCAATCACT 1793
571 YrGlyTyGluGlyThrTrpGlyProIleValTrp.....GlyThr 584
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1794 ATGGGTATCAAGGAGACTGAGAT...ATTGTTGGTCGACAGATCAACT 1840
585 GlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyIleP 601
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1841 GCAAAACAAAAAATGCTACCTTACTTGAGCTAAACAGATACAGACC 1890
601 CAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAlaP 618
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1891 GAATCCAAAGCTCAGGACCTTGTCTCAATAGCCCTGGGGCTTCTT 1940
618 heIleAspIleSerSerLeuHisTyLeuMetGluThrAlaAsnGluGly 634
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1941 TTGTGATGTCGCTCCATTCACAGCTCATGACCGGACGACCAAGTTG 1990
635 LeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPheHis 651
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1991 TTATCTGCTCAACAATTTGGGTATCAGGAATCGGAGACTTTTGCA 2040
651 sLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyTyT 668
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2041 TGAAGATCAGAAAGGAACCAAGTAGTATCTGATCTAGCGGGGCTT 2090
668 YrValIleGlyLysAsnLeuHisThrCysSerAspLysIleLeuSerAla 684
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2091 ATGCATTAGAGAGAGATTCTTCACGGCTTCGAAATTTCTTTAATTTT 2140
685 AlaPheCysGlnLeuPheGlyArgAspArgAspTyPheValAlaLys 701
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2141 GCTTTTGTGAGCTTTTGGCTACAGACACCATCTTGTGCTAAGAA 2190
701 ngInGlyThrValTyGlyThrLeuTyTrpGlnHis.....A 715
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2191 CCATACCCATGATATGAGGGCAATGATGACCGACACCTCGAGAGAT 2240
715 snGluThrTyIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2241 CTAGACCCCTGGCTAGAT.....TTGTCA 2266
732 TyrValProThrGluIleProValLeuPheSerGlyAsnLeuSerTyTh 748
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2267 GGAATTCGACTCCCTACCTTTGTCTTCAATGCTCGTTGCTTATGC 2316
748 HisThrAspAsnAspLeuLysThrLysTyThrThrTyProThrVal 765
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2317 CCATACCGCATATACATGACCAAGTACACTGCTATTCTCTGTATA 2366
765 ysGlySerTrpGlyAsnAspSerPheAlaLeuGluPheGlyArgAla 781

```

```

||||| ::::: ||||| ::::: ||||| ::::: |||||
2367 AGGAGAGCTGGGAAATGATGCTTCGCTATAGATGTGAGGAGCTATC 2416
782 ProIleCysLeuAsp...GluSerAlaLeuPheGluGlnIleTyMetProPh 797
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2417 CCGGTAGTGTGCTTACAGACGTCGGCTTGGGTGATGATCCACAGCCATT 2466
797 emetLysLeuGlnPheValTyAlaHisGlnGluGlyPheLysGluGln 814
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2467 TCTAACCTAGATGATCTATCCATCAGATCAGAAATGATTAAGGAAACG 2516
814 IYThrGluAlaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeu 830
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2517 GCACAGAAAGCGCTTCTTCCAAAGTGAAGACCTTCAATCTAGCGGCTT 2566
831 ProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAlaThrTy 847
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2567 CCTGATAGGATAAATTTGAATAATCTCCGAT.....AAGCTACGTA 2610
847 rAsnLeuThrLeuGlyTyThrValAspLeuValArgSerAsnProAspC 864
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2611 TGATCTCTCCATAGCTTACGTTCCCGATGTGATTCGTAATGATCCAGCT 2660
864 YThrThrThrLeuArgIleSerGlyAspSerTrpLysThrPheGlyThr 880
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2661 GCACGACACACTTATAGTTCTTCTGGGATTTCTGTCGACATGTGSTRCA 2710
881 AsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPh 897
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2711 AGCTTCTCTACACAAGCTCTTCTGTACGTCTCGAATCATCATGCCCTT 2760
897 eAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuArgGlyS 914
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2761 TGCTTCAACCTTGAAGTTTCACTGATGATGGAAGTCGAGTTGGAGCTT 2810
914 eSerArgAsnTyAsnValAspLeuGlyAlaLysTyArgLysPhe 928
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2811 CTCTCTGAGCTATGATCATCTTGTGAGAGAAATTCGGATT 2854

seq_name: /STD6/gcgdata/geneseq/geneseqn/NA1999.DAT.X06821
seq_documentation_block:
ID X06821 standard; DNA: 2757 BP.
XX
AC X06821:
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp9 DNA.
XX
KW Omp9; outer membrane protein 9; surface exposed protein; antigen;
  infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX
PA (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
  Mygind P;
XX
DR WPI: 1999-105610/09.
XX
DR P-PSDB: W06822.
XX
PT Species-specific test for identifying mammals infected with

```

PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins

XX Claim 6, Page 55-56; 115pp; English.

CC This DNA sequence codes for the novel 96.7 kDa surface exposed
CC protein Omp4 (see W88422) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp5 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9, and 15
CC in the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

50 Sequence 2757 BP; 770 A; 564 C; 642 G; 781 T; 0 other;

alignment_scores:

Quality: 2001.00 Length: 953
Ratio: 2.814 Gaps: 21
Percent Similarity: 74.607 Percent Identity: 45.435

alignment_block:

US-09-428-122-2 x X06821 ..

Align seq 1/1 to: X06821 from: 1 to: 2757

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
1 ATGAGATGCTCTTTCTTCTCTGTTATTAATCATCTCAAGCTTTC 50
17 OleuSerMetIle.....AlaThrGluThrValLeuAspSers 30
1 TCTCTTAATGAGTGTCTTCTGCAGATGCTGCCGATCTCAATTAGGAGTC 100
30 eAlaSerPheAspGlyAsn...LysAsnGlyAsnPheSerValArgGlu 45
101 GTGACAGATTAAATGATGATACAGACACACAGAAATTACTCTAAAGCG 150
46 SerGlnGluAspAla...GlyThrThrTyrLeuPheLysGlyAsnValh 61
151 GCACATTCTGATCTAGTGCACACACCTATATTTCTCATGGGGATGCTC 200
61 rLeuGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheA 78
201 GATAAGCCAA...GCAGGAAACAAACAGAGCTTAACACAAAGTTGTTT 247
78 snAsnThrLysGlyAspLeuThrPheThrLysGlnGlyAsnSerLeuLeu 94
248 CTAACTGAGGAAATCTTACCTTTAGGGAACGATTTCTTCAT 297
95 PheGlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSe 111
298 TTTCACATATATTTTCTGCTACTGTCAGGAGTGTGTTGTTAGCATAC 347
111 rValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheI 128
348 AGCAGCTTCTGATATACGAAATTCAGAGATTTTCAACTCTTCGAGTC 397
```

```

128 leAlaSerProGlySerSerIleThrThrGlyGlyAlaValSerCys 144
398 TTGCAGCTCTAG.....ACCACAGGTAAAGAGCATTAATAT 438
145 SerThrGlySerLeu.....SerLeuThrLysAsnValSe 156
439 ACCATGCTCTGCTGTTGTTGAGAGATATGGAGATCTTACCAAAAT..... 483
156 rLeuLeuPheSerLysAsnPheSerThrAspAsnGlyAlaIleThrA 173
484 .....GAAATGCTCTAGTGAATATGGGAGCATCATATA 520
173 lAluThrIleSerLeuThrGlyThrThrMetSerAlaLeuPheSerGlu 189
521 CGAAGCTTGTCTTTGACTGCGAGTACGGGTTTGTAGCGTCTTGCG 570
190 AsnThrSerSerLysGlyGlyAlaIleGluThrSerAspAlaLeuTh 206
571 AATGCTCTGTCGACACAGGGAGCGATATGCTTGTGCTGCTGTG 620
206 rIleThrGlyAsnGlnGlyIleValSerPheSerAspAsnThrSerSera 223
621 GATTTCTGAGAAATGCAAGGAATCTTGAGCTTGGAACAAACAGTGCACA 670
223 spSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsn 239
671 CATCAGGAGCGGATCTGCTGGAAGGAACTTGATGATCTCCATAAC 720
240 AlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSerTh 256
721 CAAATATCTTTTTCATGCTGCTCAAGCACT..... 753
256 rThrGlyAspMetSerGlyAlaIleCysAlaTyrLys.....Ths 271
754 .....ACAAATGCGGACCTTATGATTGATTAACAAGCAGGCGCA 793
271 eThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuPheSer 287
794 ACCCAGACCTATCTGCTTTCAGAAATGAGAGCTTCATTTTCG 843
288 AsnAsnThrSerThrThrAlaGlyAlaIleIleValLysLysLeuG 304
844 AATACACAGCAGAAATATGAGAGTGCATTAACCAAAATTTGTT 893
304 uLeuAlaSerGly...GlyLeuThrLeuPheSerAsnSerValAsnG 320
894 GTTATCTCAGGACGAGAGAGTGTATTTCTTAACAAAGAGCTGCGA 943
320 lGlyThrAlaProLysGlyAlaIleAlaIleGluAspSerGlyLys 336
944 ATGCTACT...CCTAAGAGAGGCAATTGCGATTCATGATTCGAGAG 990
337 LeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnTh 353
991 ATTAGATTTCGAGATCTCGCAATATCATATTTCGAGGCAATACAT 1038
353 lThrSerThrThr.....ProGlyThrAsnArgSerSerIleAspL 367
1039 .ACGAGCATATACAGAGAGCTCGAGAGTGTACACGAAATGCTATAGATC 1087
367 euGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAla 383
1088 TTGCATTCGAATCAAAATTTTAATCTCCGAGCGACTCGGGAAATATA 1137
384 lIeTyrPheTyrAspProIleThrThrLysSerThrThrValThrAs 400
1138 GTTATTTCTATGATCATATACG.....AGCTCAGAGGACTACGA 1178
400 pValLeuLysValAsnGluThrProAlaAspSerAlaLeuIleThrIrg 417
1179 TTAGCTCTCTTGAATTAACGTCAGCAGGATCTGGAATATCTTATGAG 1228
417 lYAsnIleIlePheThrGlyGlyLysLeuSerGluThrGluAlaAlaAsp 433
```


XX W09858953-A2.
 PN
 XX 30-DEC-1998.
 PD
 XX 19-JUN-1998; 98WO-DK00266.
 PF
 XX 23-JUN-1997; 97DK-0000744.
 PR
 XX (BIRK/) BIRKELUND S.
 PA (CHRIL/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX WPI: 1999-105610/09.
 DR P-PSDB: W88429.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins

PS Disclosure: Page 73-77; 115pp; English.

CC This DNA sequence encodes the novel surface exposed protein Omp5
 CC (see W88429) of Chlamydia pneumoniae, a human respiratory pathogen.
 CC It is described as a subsequence of a claimed nucleic acid fragment
 CC (see X06817) encoding Omp5 (see W88418). The invention provides a
 CC new species specific test for identifying mammals (including
 CC humans) infected with C. pneumoniae. The test comprises detecting
 CC antibodies specific for surface exposed proteins Omp4-Omp5 (see
 CC W88417-28) or detecting nucleic acid fragments encoding them (see
 CC X06816-27), especially by PCR. The proteins are also used in the
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
 CC and proteins can also be used in the immunization of mammals, the
 CC nucleic acids being particularly useful as DNA vaccines for
 CC effecting in vivo expression of antigens. The vaccines may also
 CC prevent atherosclerosis and bronchial asthma, which are possibly
 CC associated with C. pneumoniae.

SQ Sequence 3000 BP; 867 A; 597 C; 658 G; 878 T; 0 other;

alignment_scores:
 Quality: 2000.00 Length: 936
 Ratio: 2.801 Gaps: 20
 Percent Similarity: 76.282 Percent Identity: 46.261

alignment_block:
 US-09-428-122-2 x X06828 ..

Align seg 1/1 to: X06828 from: 1 to: 3000

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePh 16
 259 ATGAATATCGCAATTTTCCTGCTAGTCTCTCTCGACATTTGGCATTTT 308
 16 e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSer 30
 309 TACTAGTTGTTCCACTGTTTTCGTCGACAGCTGAATAATAGGCCCT 358
 30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgLys 45
 359 CTGATACCTTTGACGAGAGACTACACAGCAGCTTACTCTCTTAATAAT 408
 46 SerGlnGluAspAlaGlyThrThrTyLeuPheLysGlyAsnValThrLe 62
 409 ACG.....ACTACTGGAATAGACTATACCTCTACAGAGATATTACTCT 452
 62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
 453 GCAAAACCTT...GGGATTCGGCAGCTTACAGAAAGGTTGTTCTG 499

79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
 500 ACACACGCAATCTTAAGCTTTGCGCGTAAGGGGATCTCACTTTCTTTT 549
 96 GlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
 550 TTAATATTAACTCT...AGTCTGAAGGCGCAGCAGCTT...TCTGTAC 593
 112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
 594 AACTGATTAATAATCTGCGCTACACAGCATTTTCAGCTTACTTTCTTAG 643
 129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
 644 CGGCCCATCATCGTAAATCAACCCCTCAGAGAAAGGTGACGTTAA 693
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
 694 TGTGGA...GGGATCTTACATTGATACAAATGACATTTTATTATTA 740
 741 ACAAGATTACTGTGAGGAAATGCGGAGCCATTTACCAAGATCTTT 790
 160 rLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeu 177
 177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSer 193
 791 CTTTGAAATACAGACGCGGATTCGATTTCTTGAAGGATTAATATGAC 840
 194LysLysGlyAlaIleGlnThrSerAspAlaLeuThrI 207
 841 GCAACAGGAAATAGGTGGGCTATTGCTACGACTGACTGATATAT 890
 207 eThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerAsp 224
 891 TACAATAATACGCGCTCCACCTCTTCGAAACATATGCTGAAGCTG 940
 224 erGlyAlaAlaIlePheThrGluAlaSerValThrLeuSerAsnAla 240
 941 CAGGTGAGCTATTAATACACAGGAACTGACATTTACAGGAAATGAC 990
 241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 257
 991 TCTCTGTATTCTGAAATAGTGACAA.....GCGACCGC 1028
 257 rGlyAspMetSerGlyGlyAlaIleLysAlaIleThrSerThrAsp 274
 1029 AGGA.....AATGAGAGCTCTT.....TCTGAGATG 1057
 274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
 1058 CCGATGTACCATATCTCGAATCAGAGTCACTTTCTCAGAAACCA 1107
 291 SerThrThrAlaGlyGlyAlaIleThrValLysLysLeuGluLeuAla 307
 1108 GCGTAGCTAATAGCGGAGCCATTATGCTTAAGAGCTTACAGCTGCTC 1157
 307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGly 322
 1158 CGGGGGGGGGGGGATCTCTTTTCTAACAATATATCCAAAGGTACCA 1207
 322 hrAlaProLysGlyAlaIleAlaIleGluAspSerGlyGluLeuSer 338
 1208 CTGACAGTATGCTGAGCCATTCTATACTGCGCAGCTGAGAGTGTGT 1257
 339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 355
 1258 CTTTAGAGAGAGCAGGGGACATTACTTCAATGGGAATGCAATGTTGC 1307
 355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSer 371
 1308 AACTACACCACTACAAAGAAATTTATTTGACATAGATGACTCTCTG 1357

```

371  laryethrhlaleuagsserlalaaglyarAlalephrhyr 387
1358  CAAGATCCAGCAATTACGTGCAATATCTGGGCATAGCACTCTTTTCTAC 1407
388  ASPProIleThrThyGlySerSerThrThValIthAspValLeuIysVa 404
1408  GATCGATTAAGTGTAAATCGGCTGGGATCTTCAGATACTTTAAATCT 1457
404  lAsnGluThrProAlaAspSerAlaLeuGlnItyrThrGlyAsnIleIleP 421
1458  CAAATAGGCGTGAATGCAGTAATAGAACATTAATATGCTGGCTGATGTTT 1507
421  heThrGlyIuIysLeuSerGluThrGlnAlaAlaAspSerLysAsnLeu 437
1508  TTTCTGCTGAAGAAGCTCTGCAAGATGAAGCAAAAGTTGCACACAACTTC 1557
438  ThrSerLysLeuLeuGlnInProValIthrLeuSerGlyGlyThrLeuSerLe 454
1558  ACTTGTACCTCCATAGAGCGCTGTAACTCACTCAATCGCAGGAAATTTGTACT 1607
454  uIysrHsIsgValAlThrLeuGlnIthGlnAlaPheThrGlnAlaAspS 471
1608  TAAAGCTGGTGTCACCTCTCGATACGAAAGCTTTTACTCAGACCGCGGTT 1657
471  eTArgLeuGlnuMetAspValGlyThrThrLeuGluProAla...AspThr 486
1658  CCTCTGTATTATGTATGATCGCGGCACACACTTAAAGCAAGTACAGAGAGC 1707
487  SerThrIAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaIly 503
1708  GTCACCTTTACAGGCTTTCCATCTCTGTAGACCTTTTAGCGGAGGATTA 1755
503  slYsAlaIyIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
1758  GAAAGTGTAAATGGCTGCTTCGACCAAGTAAATGATGAGCCCTTATGTG 1807
520  lYThrIleThrLeuLeuAspProIthGlyThrPheTyGlnAsnHisSer 536
1808  GTCCGATTTCTTTTGGATACCAAGAGGAATGCTTATGAAATACAGAG 1857
537  LeuArgAsnProGlnSerIleAspIleLeuGlnuLeuLysAlaSerGlyTh 553
1858  TTAGCAAAAACGACAGACTTTTCATTTGGACGCTGTGCTGCTGGTATAC 1907
553  rValThrSerThrAlaValIThrProAspProIleMetGlyLysPheN 570
1908  TGCACAACTACAGATGTTCACGCGCTTCTCACTCACTCAACTCTTACGC 1957
570  lstyGlyTyrglnGluItyrThrArgIleProIleValItrP... 582
1958  ACTATGGGTATCAAGGTACTTGGGA...ATGACTTGGGTGATGATACCC 2004
583  ..GlyTyGlyAlaSerThrIthrAlaThrPheAsnItrThrLysThGly 598
2005  GCACAGCACTCCAAAGCACTAAAGCAGACGACATTAAGCTTGACCAATACAG 2054
598  uTyxIleIleProAsnProGlnArgIleGlySerLeuValProAsnSerIleuT 615
2055  CTATCTTCGGAATCTGCAGCGGACGCGCAAGAGACTTTTAACTGCTTAATACCTTT 2104
615  rPAsnAlaPheIleAspIleSerSerLeuHisTygLeuMetGluIthrAla 631
2105  GGGGATCTTTTTCAGACATCCAAAGGATTCAAAGGTGTCAATAGAGAGAGT 2154
632  AsnGluGlyLeuGlnGluLysArgAlaPheTrpCysAlaGlyLeuSerAs 648
2155  GCTTTGACSTCTTTGTCACATGAGAGCTTCTGGGTGGCGAGGTGCGCA 2204
648  nPhePheHisLysAspSerThrLysThyTrpArgLysPheArgHisLys 665
2205  TTTCTTAAGTAAAGATAAAGAAAGGGGAAAGCAAGATATACCTGTAAAT 2254
665  eTcIyGlyItyrValIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681

```

2255	CTGGTGAATATGCTATCGGAGGCGACCGAAACTTGTCTGAAAACCTTA	2304
682	LeuSerAlaIaIaPheCysGlnLeuPheGlyArgAspArgAspTyrPheA	698
2305	ATTAGCTTTGCCCTTTGGCAACCTTTGGTAGCATAAAGATTCTTAGT	2354
698	IAlaIysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrLinhSA	715
2355	CGCTAAAAACATCACTGATCATTAAGCAGGAGCCCTTGATATCCAAAC	2403
715	snluthrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer	731
2404ATTACGAATGTATAGTGGGTGATAGATGGTTGCTCTTAGAT	2442
732	TyrValPro.....ThrGluIleProValLeuPheSerGlyAsnLe	745
2443	AAACTTCCTGGCTCTTGGAGCTCAAAACCCCTCTTTAGAAAGGCACT	2492
745	userTyrThrIleSthAspAsnAspLeuLysThrLysTyrThrTyrP	762
2493	CGCTTTATAGCAGCTCGATATATCTGAAAGCAAAATATATCTCGTATC	2542
762	rothrValLysGlySerTyrGlyAsnAspSerPheAlaLeuGluPheGly	778
2543	CTAGAGGGAAGGTTCTTGGGGGATATATCTTTAACTATGATGTGGGA	2592
779	GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTyrMe	795
2593	GCTTCTTCTATCTTTCCTGGAATACCTGCATGTGTTTGATACCTATGC	2642
795	trProPheMetLysLeuGlnPheValTyrAlaIleHisGlnGluPheLysG	812
2643	TCCATCACTCAAACTGATATCTGACCTTATAACGTCAGGACAGCTTCGCG	2692
812	IuGlnGlyThrGluAlaArgGluPheLysSerArgLeuValAsnLeu	828
2693	AGAAAGGTACAGAGGAGGAAAGTCTTTTATGACAGCAACCTTCAATTTA	2742
829	AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAl	845
2743	TCTTGGCTATAGGGGGTGAAGTTTGAGAAAGTCTCTGATTTGATGACTT	2792
845	atThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuValArgSerAsnP	862
2793	TTCTTATGACTACTTATCTATCTATGTCTGTGACTTATCCGCATGATGC	2842
862	roAspCysThrThrThrLeuArgIleSerGlyAspSerTyrLysThrPhe	878
2843	CCAAATGCACTACAGCAGCTGTATATCAGCGAGGAGCTCTTGGGAACCTAT	2892
879	GlyThrAsnLeuAlaArgGlnAlaLeuValLeuAlaIleGlyAsnHisPh	895
2893	GCCAAATAACTTACACACAGCGCTTTCAGATGCGTCAGCAGCAGCTCACTA	2942
895	ecySPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuA	912
2943	CGCCTTCTCTCCATGTTTAAAGTGTCTGGCGAGTTGTCTTTGAAGTTC	2992
912	rgGlySer 914	
2993	GTGAGTCC 3000	
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA199.DAT.X06822		
seq_documentation_block:		
ID	X06822 standard; DNA. 2787 BP.	
AC	X06822:	
XX		
DT	26-APR-1999 (first entry)	
XX		
DE	Chlamydia pneumoniae surface exposed protein Omp10 DNA.	

XX Omp10; outer membrane protein 10; surface exposed protein; antigen;
 KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
 XX Chlamydia pneumoniae.

XX WO9858953-A2.

XX 30-DEC-1998.

XX 19-JUN-1998; 98WO-DK00266.

XX 23-JUN-1997; 97DK-0000744.

XX (BIRK/) BIRKELUND S.
 XX (CHR1/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;

XX WPI: 1999-105610/09.
 XX P-PSDB: W88423.

PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins

PS Claim 6; Page 59; 115pp: English.

XX This DNA sequence codes for the novel 98.4 kDa surface exposed
 CC protein Omp10 (see W88423) of the human respiratory pathogen
 CC Chlamydia pneumoniae. By generating antibodies against C.
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
 CC was obtained which reacted with outer membrane proteins. The
 CC antibody was used to identify the genes (see X06816-27) encoding
 CC Omp4-omp5 proteins (see W88417-28) in an expression library of
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
 CC the other, and encode polypeptides of about 89.6-100.3 kDa and
 CC about 56.1 kDa. The invention provides a new species specific test
 CC for identifying mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-omp5 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.

XX Sequence 2787 BP: 815 A; 689 C; 535 G; 748 T; 0 other;

alignment_scores:

Quality: 1986.00 Length: 950
 Ratio: 2.770 Gaps: 19
 Percent similarity: 75.474 Percent identity: 45.263

alignment_block:

US-09-428-122-2 x X06822 ..

Align seg 1/1 to: X06822 from: 1 to: 2787

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
 |||||
 1 ATGAATCTCTCTTCATGTTTGTATCTCGTCACTTTAGCAGCTTC 50
 17 OleuSerMet.....IleAlaThrGluThrValLeuA 28
 |||||
 51 CTTGCTACTAATATTCTCTGCGTTGCTGCTGTGTTGTAATCAATCTAG 100

28 sPseSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerValArg 44
 : : : : :
 101 GACCTACCAATAGCTTCTCTGGA.....CCAGCAACTACCTCTCCCA 144
 45 GluSerGlnGluAspAla...GlyThrThrLeuPheLysGlyAsnVa 60
 : : : : :
 145 GCCCAACACAAATGACAGATGAGACTATCTAATATCTAACAGGGAGCT 194
 60 lThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysP 77
 : : : : :
 195 CTCATCTACCAAT...GCAGATCTCCGACAGCTCTTACCCCTTCTCCT 241
 77 heAsnAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeu 93
 : : : : :
 242 TTAAGCAACTCTCTGGAATCTTTCTTCCAGGCGCAGCTTACCATT 291
 94 LeuPheGlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSe 110
 : : : : :
 292 CTCTACCAAAATATCGATCGGGA.....GCGAAGCTACCTTTACCA 335
 110 rSerValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerP 127
 : : : : :
 336 TACAGCTCCAAATAGCTTCTCTCTTCAGGATTCCTATTGTCTAC 385
 127 heIleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSer 143
 : : : : :
 386 TAATTCANACC.....ACGAATGCTACACAGACAGAGCCATCAG 429
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
 : : : : :
 430 ...TCCACAGAGGCTGTGTTTATTTCAGTCGCAACTAAGTTGCTACTTTGG 476
 160 rLysAsnPheSerThrPheAsnGlyGlyAlaIleThrAlaLysThrLys 177
 : : : : :
 477 CCAAACTTTCTTAATGACATGGAGGCGCCCTCCAGGCGCTCTATCA 526
 177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
 : : : : :
 527 GCTATATCG...CTAAACCCCACTAAGCTTTGCCAAACCAAGCAACG 573
 194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAs 210
 : : : : :
 574 CAAAGGGGGGTGCCCTTATTCACGCGGAGGATTCAAATTAACAATAC 623
 210 nGlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAla 227
 : : : : :
 624 GTTAACACACACATCATTTCTGAAATACCGCGCCGCAACAATGGCGAG 673
 227 lAllePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSer 243
 : : : : :
 674 CCATTTCACGAGAGCTAGCAGTTTATTAGACGACCAACAAGCAATTAGC 723
 244 PheIleAspAsnLysValThrGlyAlaSerSerSerThrThrGlyAspMe 260
 : : : : :
 724 TTATTAACCAATAGTGTGACCGCAACTCAGCT..... 756
 260 tSerGlyGlyAlaIleCysAlaThrLysThrSerThrAspThrLysVal. 276
 : : : : :
 757 .ACAGGGGAGCCATTTAGCTGATAGTATACATCAGCCCCCAACCACTGT 805
 277 ..ThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnAsnThrSerThr 292
 : : : : :
 806 TAACCTATATCAGACACAGGGGAACTGAACTTATAGGAATATACAGCAT 855
 293 ThrAlaGlyGlyAlaIleGlyValLysLysLeuGluLeuAlaSerGlyAl 309
 : : : : :
 856 ACTAGTGTGGGGGAGTTATACGACAAATCTAGTTCTTTCTTCTGAGG 905
 309 yLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysG 326
 : : : : :
 906 ACCTAGCCTTTTAAACAAACTGTGTATAGATACGACGCTCCCTTAG 955
 326 LylGlyAlaIleAlaIleGluAspSerGlyGluLeuSerLeuSerAlaAsp 342


```

956 GAGGAGCAATTCGATTCGACTCTGGATCTTGTAGTCTTGGCTGT 1005
343 SerGlyAspIleValPheLeuGlyAsnThrVal.....ThSerTh 356
1006 GGTGAGACATCAGCTTTGGAAGGAACACAGTACGAAGAGCTTCTTC 1055
356 rThProGlyThrAsnArgSerSerIleAspLeuGly...ThSerAlaL 372
1056 GAGTCAGACCACTACAGAAATTCATTACATCGAAACACCAATGCTA 1105
372 ySmetThrAlaLeuArgSerAlaIaIaGlyArgAlaIleTyrPheTyrSp 388
1106 AGATTGTACAGCTGGAGCTCTCTCAGGCAATACATATCTACTCTATG 1155
389 ProIleThrThrGlySerSerThrThrValThrAspValLeuLysValAs 405
1156 CCTATAACACTAACCATGCTGACAGCTCTCTCAGATGCTTAAGCTTAA 1205
405 nGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIlePheT 422
1206 TGGTCTTGACCTTGCAGGGAATCCTGCATATCAAGAACCATCGATT 1255
422 hrGlyGlyLysLeuSerGlnThrGlnAlaAlaAspSerLysAsnLeuThr 438
1256 CTGAGAGAACACTCTCGAGACAGAGCTGCAGAGCTGATATCTCAAA 1305
439 SerLysLeuLeuGlnProValThrLeuSerGlyThrLeuSerLeuLys 455
1306 TCTCAATTCAGCAACCTCTTAACCTCTTCCGGAGGCAACTCTCTTAA 1355
455 sHISgLyValThrLeuGlnThrGlnAlaPheThrGlnGlnAlaAspSer 472
1356 ATCAAGAGTCACTAGTGTCTAGTCTTTCGCAATCTCCGGCTCTTA 1405
472 rGLeuGlnMetAspValGlyThrThrLeuGlnProAlaAspThrSerThr 488
1406 CCTCTCTGATGATGACGAGGACACATAGAAACCGCTGATGGATCCT 1455
489 IleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLysLysAl 505
1456 ATCAATATCTTGTCTCAATGTAGATCTTAAGAGACCAAGAACAGGC 1505
505 aLysIleGlnThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrI 522
1506 TACGCTAAACACACACAGCAAGTCAGACAGTCACTTATCTGGATGTC 1555
522 IeThrIleLeuAspProThrLeuThrPheTyrGluAsnHisSerLeuArg 538
1556 TCTCTCTGTAGATCTCTGGAATGCTCTAGAAAGATGCTCTGTGAAAT 1605
539 AsnProGlnSerTyrAspIleLeuGlnLeuLysAlaSer.....GlyTh 553
1606 AACCTCTCAGCTTTTCTCTCTCACTCTACTCTGAGACCCCGGCA 1655
553 rValThrSerThrAlaValThrProAspProIleMetGlyLysLysPheN 570
1656 TATTCACATCACAGACTTACCTGCTGATCCCTTAGAAAAAATCTTAAT 1705
570 IeTyrGlyTyrGlnGlyThrTyrPheLysProIleValThrGlyThrGlyAla 586
1706 ATTGGGATACCAAGGAATGGGCA...TATCTTGGCAGAGGATCT 1752
587 SerThr.....ThrAlaThrPheAsnThrPheThrLysThrGlyTyrI 600
1753 GCGCATTAATCCAAAGCAGCTTCTTACCTGACAAAAACAGGATGAA 1802
600 eProAsnProGlnArgIleGlySerLeuValProAsnSerLeuTyrAsn 617
1803 TCCGATTCCTAGCGCTGCGTGAACCTTAGTGTCAACAGCGTATGGGAT 1852
617 IapelleAspIleSerSerLeuHisTyrLeuMetGlnThrAlaAsnGln 633
:::
1853 CCTTTGTATGTGCGCTCCATACACAGCTTGTAGCCACTAAAGTACGC 1902
634 GlyLeuGlnGlyAspArgAlaPheThrCysAlaGlyLeuSerAsnPheP 650
1903 CAATCTCAAGAACTCGCGGATCTGGTGTGAAGGATCTCGAACTTCT 1952
650 eHisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
1953 CCATAAAGATAGCAGCAAGATAAATTAAGGTTTCCGACATAGTGCAG 2002
667 LyTyrValIleGlyLysLeuHisThrCysSerAspLysIleLeuSer 683
2003 GTTATGTGTAGAGGACACTAACACATTAGCTTGTGTAATCTTATGACT 2052
684 AlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAla 700
2053 GCACCTTCTGCCAATATTTCGGAAGATAGAGATCATTATTAATTA 2102
700 sAsnGlnGlyThrValTyrGlyThrLeuTyrTyrGlnHisAsnGln 717
2103 AAATAGAGCTTGTGCTTAGAGCTTCTCTCACTCTCACCACATCTAGCG 2152
717 hrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrVal 733
2153 CCTGTCTCT.....CCAGCTGTATAGCTTACT 2184
734 Pro.....ThrGlnIleProValLeuPheSerGlyAsnLeuSerTyr 747
2185 CCTGATCTGAAGTGAAGTGAAGCCCTCTCTGTGATCTCGATCACACTA 2234
747 rThrHisThrAspAsnLeuLysThrLysTyrThrThrTyrProThrV 764
2235 TATCTATAGTAAATCTATGAAGACCTATTACACCCAAAGCACCAGG 2284
764 aLysGlySerTyrGlyAsnAspSerPheAlaLeuGlnPheGlyArg 780
2285 GAGAGACTCTGGTATATAGCGGTTGCGCTGTGGAATCTGGAGCTCC 2334
781 AlaPro...IleCysLeuAspGlnSerAlaLeuPheGlnGlnTyrMetPr 796
2335 CTACACACACACTGTTTAAGCAGAGGGTCTCTCCACCGATTTTTC 2384
796 oPheMetLysLeuGlnPheValTyrAlaHisGlnGlyPheLysGln 813
2385 TTTTCATTAAGTAGAAGCTTCGTACATACACCAAGATAGCTTCAAGAAC 2434
813 InGlyThrGln...AlaArgGlnPheGlySerSerArgLeuValAsnLeu 828
2435 GTATATCTACTCTGTGATGATCTTTCGATGAGCGGTATTAATTAAGCTC 2484
829 AlaLeuProIleGlyIleArgPheAspLysGlnSerAspCysGlnAspAl 845
2485 TCTGTGCTATTTGAATTTACCTTCGAGAGATTCCTCGAAGAACGCTGC 2534
845 aThrTyrAsnLeuThrLeuGlnTyrThrValAspLeuValAspSerAsn 862
2535 GTCTTAGAGAGCTACTGTACATACGTGGCAGATGTACGAAGAATC 2584
862 roAspCysThrThrThrLeuArgIleSerGlyAspSerTyrPheThrPhe 878
2585 CTGACTGCAGACAGCTCTCTTAACAACATCTCTGTGGAAGAACTACA 2634
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisP 895
2635 GGAAGCAATCTCAAGACAGAGCTGATGGAAGACAGAGGATCTTTTA 2684
895 eCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSerPheGlnLeu 912
2685 TGCTTCTCTCCAAATCTGAGGTCAACACTATCTATGAATTC 2734
912 rGlyLysSerArgAsnTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
2735 GTGATCTTCACGACCTACAAATGCAGATCTTGAGGTAGTTCAGTTTC 2784

```

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA1999.DAT.X06823

seq_documentation_block:

ID X06823 standard; DNA; 2793 BP.

AC X06823;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp11 DNA.

KM Omp11; Outer membrane protein 11; surface exposed protein; antigen;

XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygland P;

DR WPI: 1999-105610/09.
DR P-PSDB: W88424.

Species-specific test for identifying mammals infected with
Chlamydia pneumoniae - comprises detecting antibodies specific for
outer membrane proteins of C. pneumoniae or nucleic acids encoding
these proteins

Claim 6; Page 62-63; 115pp; English.

This DNA sequence codes for the novel 97.6 kDa surface exposed
protein Omp11 (see W88424) of the human respiratory pathogen
Chlamydia pneumoniae. By generating antibodies against C.
pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
was obtained which reacted with outer membrane proteins. The
Omp4-Omp15 proteins (see W88417-28) in an expression library of
C. pneumoniae DNA. The genes are situated in 2 gene clusters:
Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
the other, and encode polypeptides of about 89,6-100.3 kDa and
about 56.1 kDa. The invention provides a new species specific test
for identifying mammals (including humans) infected with Chlamydia
pneumoniae. The test comprises detecting antibodies specific for
Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
membrane proteins, especially by PCR. The proteins are also used
in the diagnosis of C. pneumoniae infection in mammals. The
nucleic acids and proteins can also be used in the immunization of
mammals, the nucleic acids being particularly useful as DNA
vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

Sequence 2793 BP; 803 A; 681 C; 567 G; 742 T; 0 other;

alignment_scores:

Quality: 1927.00 Length: 953
Ratio: 2.706 Gaps: 17
Percent Similarity: 74.711 Percent Identity: 43.442

alignment_block:

US-09-428-122-2 x X06823

Align seg 1/1 to: X06823 from: 1 to: 2793

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
  |||||
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 ATGAAATACCTTGCACAAACCTCTGATCTTCGACTCTTGCTACTCC 50
17 GLeuSerMet...IleAlaThr.....GluThrValLeuAspSer 30
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 CATTTATTTAGCATGTGCACTTACGAGCAGATGCTTTTATCCCTTA 100
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 eAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerValArgGluSer 46
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 CAGATAGCTTTGATGAGCGCGCGCTCTACATTCTCCAAATCTACA 150
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47 GlnGluAspAlaGlyThrThrThrTyLeuPheGlyAsnValThrLeu 63
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 GCAGATGCCAATGGACGAACTATGCTTATCAGAAATGCTATATA.. 198
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 uAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnSer 80
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199 .AACGATGCTGGGAAAGGCACACATTACAGCTGCTCTTTACAGAA 247
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 hrlYsgIAspLeuThrPheThrGlyAsnGlyAsnSerLeuPheGln 96
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 CTACGGGTGATCTGACATTTACTGAAAGGATCTCAATTTTCATTCAAC 297
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97 ThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerValVa 113
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 ACGTAGATGGGGTTCGAATGCAAGTCAAGAGCTGGCA...AGCACACTGC 344
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 LAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleAla 130
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 TGATTAAGCCCTTACATTCACAGATTTCTAACCTTCTTCAATGGACG 394
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 ePProGlySerSerIleThrThrGlyLysGlyAlaValSerCysSerThr 146
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
395 CTCCTGAACTACAGTGTCTCAGGAAATAACTCTTAG...TCGCA 441
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 GlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSerLysAsnPh 163
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 GAGGCTTAATCTTACCGATATGAAACGATCTCTTAGCCAAACGT 491
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 eSerThrAsp.....AsnGlyGlyAlaIleThrAlaLysThrLeu 177
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 CTCATGATACGATATAACATGCGGAGCATCCACCAAAACTCTTT 541
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 CTATTCTGGGAATACCTCTTCTAATACCTTCACTAGTATATACGCAAA 591
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrLys 210
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 AAATTAGGTGAGGAGCATCTAGCTGCGGCTGCAATGATTTTCAGAAA 641
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 nGlnGlyGluValSerPheSerAspAsnThrSerAspSerGlyAla 227
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
642 CACGCGCACTTACTCTTATGAAATATAAGAGAAACTGGGGGGGGG 691
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 lAllePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSer 243
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
692 CTCGGGCTTTGAAGCAGCTCTCGATTACTCAAAATAGCTCCCTTTTC 741
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 PheIleAspAsnLysValThrGlyAlaSerSerThrThrGlyAspMe 260
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
742 TTCTCTGAAACACTGCACACAGATGCTGCAGC..... 774
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 tSerGlyGlyAlaIleCysAlaThrLysThrSerThrAspThrLysVal 277
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
775 .AAGGCGGGGCAATTTATTTGAAAAACAGAGAGACTCTCTACTCTTA 823
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
277 hrlLeuThrGlyAsnGlnMetLeuLeuPheSerAsnAsnThrSerThr 293
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

824 CTATCTGGAAATAAAGTCTGACCTTCGCCGAGAACTCTTCACTACT 873
284 Alaelglylalailetyrvallysleuglualeualaaserglyyla 310
|||||
874 CAAGCGGAGCAATCTGTGCTAGATCTTCCGCTGCGGCC 923
310 uThleupserArasnerValasnGlylThralaProlysgly 327
|||||
924 TACCTATTTCAAATAATAGATGCGGAGCAAGCTCGAGGAGGCG 973
327 lYalailealailegluasPserglyluleuSerleuSerAlaPser 343
|||||
974 GCGCTATTGCAATTCGCACTCTGATCTTAAGTCTCTCGAAATGAA 1023
344 GlyaspIleValPheleuglyAsnThrValThrSerThr...Progl 359
|||||
1024 GGAGCAATCAGGTTCTTGGCAACACTTAACCTCAACCTCCGCCAAC 1073
359 YThrAsnArserSerIleaspIleuglyThrSerAlaIleMetHra 376
|||||
1074 ATCGACAGGAAATGCTATCTACTGGATGCTCGACAAAATTACGA 1123
376 euArgerSerAlaIleAlaIleAlaIleThyPheTyAspProle... 390
|||||
1124 TAAGGCGAGCCCAAGCCAACTATCTATCTATGATCCGATTCGAT 1173
391 ...ThrThrGlySerSerThrThrValThrAspValleuLysVal 406
|||||
1174 AACACCCAGGAGCTCA.....GACGTCGACCAATCAACCA 1211
406 uThrProAlaSPserAlaIleuglyThrGlyAsnIleIlePheTh 423
|||||
1212 ACCGATAGCAACCTGCTTATGATTAATTCAGAAAGATTTATTTCT 1261
423 lYgluLysLeuSerGlyThrcluaIalaSPserLysAsnleuThSer 439
|||||
1262 GGGAAAGCTCTCGACATGACAAAGCTCGATTAACCTTCACTACT 1311
440 LysleuLeugInProValThrLeuSerGlyThrleuSerleuLysH 456
|||||
1312 ATATTAAAGCAACATTGCTCTAGCTCTGAACTTACGACTCAAAAG 1361
456 sglValThrleugInThrgInAlaPheThGlnAlaSPserArGL 473
|||||
1362 AAATGTCAGATTAGATGCAATGTTTCACACAGCTGAAGCTCTAC 1411
473 euGluMetAspValGlyThrThrleugluProAlaSPThSerThr 489
|||||
1412 TCTCATGCAACCAAGCAAAAGCTCAAA...GCAGATCTGAAGCTAT 1458
490 Asn.....AsnleuValIleAsnIleSerSerIleaspGlyAla 504
|||||
1459 AGCTTACCAACTTGTGCTGATCTTCTGCTTAGAGGAATAAGG 1508
504 sAlaLysIleGluthrLysAlaThrSerLysAsnleuThrleuSer 521
|||||
1509 TGTGTCAATTGAACAGCAGGAGCAACAAATATACTTACACCTGC 1558
521 hrIleThrleuAspProThrGlyThrPheTyGluAsnHISerleu 537
|||||
1559 CTCTTGTTTCCAGATAGTAGCGCAATTTTATGAAGCATACGATA 1608
538 ArGAsnProGInSerTyAspIleleugluLeuLysAlaSer..... 551
|||||
1609 AACCAAGCTTACAGCAAGCTTGGTATTCACCTGCTACTGCTGC 1658
552GlyThrValThrSerThrAlaValThrProA 562
|||||
1659 TAGGATATTATATGATGCTGCTTCTACCTTCCAGTACAACTCCAG 1708
562 sPProIleMetGlyLysPheHISerTyGlyGlnGlyThrThrIle 578
|||||
1709 AACCT.....CATACGGGATACAGGACATTGGGA 1740

579 ProIleValIlePglThrGlyAlaSerThrThraIaThrPheAsnTrp 595
|||||
1741 GCCACTTGGGAGACATCAACGCAAAATCAGAGACTGATGCTGGT 1790
595 rLysThrGlyThrIleProAsnProGluArgIleGlySerleuValPro 612
|||||
1791 AACTACGGGCTACACCACTTAATCTGAGCGTAGAGCTTCGTAATCC 1840
612 snSerleuTrpAsnAlaPheIleaspIleSerSerleuHISerleuMet 628
|||||
1841 ATCATATTAGGCAATCTTACTAGCATTCGACACTACAGAGATCATG 1890
629 GluthrAlaAsnGluGlyleuGlnGlyAspAlaPheTrpCysAla 645
|||||
1891 ACATCTCAAGCGAATAGTATCTATCAGCAACAGGAGCTCGGCATAG 1940
645 YLeuSerAsnPhePheHISerLysAspSerThrLysThrArgGlyPhe 662
|||||
1941 AACTGCCAATTTCTTCATAGGATTAATCAGAACTACCAAGCATTC 1990
662 rGHisLeuSerGlyTyThrValIleGlyLysAsnleuHISerCys 678
|||||
1991 GACATAAAGCTACGCTATATGTTGGAGAAAGTCTGAAGATTTTCT 2040
679 AspLysIleLeuSerAlaIalaPheCysGlnleuPheGlyArgAsp 695
|||||
2041 GAAAAATCTTCAAGTAGCTTCTGCACTCTGCGTAAATAAAGA 2090
695 PTyPheValAlaLysAsnGlnGlyThrValTYrGlyThrleuTy 712
|||||
2091 CCGTTTATATGTTGAAATACCTCTCATACTATTAGCTCCCTATACC 2140
712 YrGlnHISAsnGluThrThrIleSerleuProCysLysleuArgPro 728
|||||
2141 TGCACATTCAGCAATTCCTAGGAGACTTCC.....ATGCCCTCA 2181
729SerleuSerTyValProThrGluIleProValleuPheSer 743
|||||
2182 TTGGAAGTATCACGCAATGCTGAAGATATTCCTCTCATTTGAAAT 2231
743 YAsnLeuSerTyThrHISThrAspAsnleuLysThrLysTyThr 760
|||||
2232 CCAAGCTACGCTACAGCTCACTAAATGATATGATACCTGCTATCT 2281
760 hrTyProThrValLysGlySerThrGlyAsnAspPheAlaIleGlu 776
|||||
2282 CTATCTGAAGCTCAAGGTTCTTGACCAATATCTGGGCTCTAGAG 2331
777 PheGlyLysArgAlaProIleCysLeu...AspGluSerAlaIlePhe 792
|||||
2332 CTGCGAGATCTGCTGCTATATCTCCCTAAGAACACACCTTCTTCCA 2381
792 uGlnTyMetProPheMetLysleuGlnPheValTYrAlaHISGlnGlu 809
|||||
2382 GGGATATTTCCTCTTAAAGTTCCAGCGAGCTTACAGCCGCCAACAA 2431
809 LyPheLysGlnGlnGlyThrGluAlaArgLysPheGlySerSerAla 825
|||||
2432 ACTTAAAGAGAGCGCGCTGAAGCCGCTCTTTGATGATGGAGACTA 2481
826 ValAsnleuAlaIleuProIleGlyIleArgPheAspLysIleSerAs 842
|||||
2482 GTGAAGTGCCTATCCCTGCGCATTCGTTGAAAAATATCCGAGAA 2531
842 sGlnAspAlaThrTyAsnleuThrleuglyTyThrValAspLeuVal 859
|||||
2532 TGAATAAATAATTTGAGATTTCTTAGCAACATTTGATGATGATAC 2581
859 rGSerAsnProAspCysThrThrThrleuArgIleSerGlyAspSerTr 875
|||||
2582 GTAAATAATCCCGTTCGCTACTTCTTAATGCTCAGTGGAGCTCTGG 2631

```

876 LysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaG1 892
      ::::: |||||
2632 ACTTGCGTATGTAAGAACCTCGCAGACAGAGCTCTTACGACAGGCTCG 2681
      ::::: |||||
892 YasnHisPheCysPheAsnSerAsnPhGluAlaPheSerGlnPheSerp 909
      ::::: |||||
2682 AAGCATCTGACTCTCTCCCATGATGTAAGACTCTCTGGGAGAGCTGCTT 2731
      ::::: |||||
909 heGluLeuArgLysSerSerArgAsnTyrAsnValAspLeuGlyAlaLys 925
      ::::: |||||
2732 ATGACCTCGTCGCTCAGACACATCTCAAAATGATGATTGGGCTAGA 2781
      ::::: |||||
926 TyrGlnPhe 928
      ::::: |||||
2782 TACTCATTC 2790
      ::::: |||||

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X06816

seq_documentation_block:
ID X06816 standard; DNA: 3200 BP.
XX
XX X06816;
XX
AC
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp4 DNA.
XX
KW Omp4; outer membrane protein 4; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; aschma; ss.
OS Chlamydia pneumoniae.
XX
XX
FH key Location/Qualifiers
FH CDS 205..2991
FH FT /*tag= a
XX
XX MO9685953-A2.
XX
PD 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mygind P;
XX WPI: 1999-105610/09.
XX P-PSDB: W88417.
XX
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 35-40; 115pp; English.
XX
XX This DNA sequence codes for the novel 98.9 kDa surface exposed
XX protein Omp4 (see W88417) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06816-27) encoding
XX Omp4-Omp15 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for

```

```

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
XX
SQ Sequence 3200 BP; 946 A; 708 C; 623 G; 923 T; 0 other;

alignment_scores:
Quality: 1855.00 Length: 949
Ratio: 2.650 Gaps: 20
Percent Similarity: 73.762 Percent Identity: 42.677

alignment_block:
US-09-428-122-2 x X06816 ..

Align seg 1/1 to: X06816 from: 1 to: 3200

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 ATGAGACTTCGATTCCTCGGCTTATGTTCTCCTCGCTTACGCTTTC 254
17 O.....LeuSerMetIleAlaThrGluThrValLeuAspSerSerAlas 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 ATGTCACTTACAGTCACTAGTACGAGAGAGACTTTATACCTGATGARA 304
32 rPheAspGlyAsn...LysAsnGlyAsnPhSerValArgLysGln 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
305 GCTTATATGGAATATCGATTCAGAGACTTACTCCAAACTCA... 351
48 GluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAs 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 .....GCCACACATATTCCTAACGAGAGATGCTCTTTACGA 392
64 nileProGlyThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrL 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
393 G...CTTGAAAGAGGCACTCCCTTATCTGACAGTGTGTTTAAAGCAACCA 439
81 yscLysAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThr 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
440 CGACATCTTACTCTTGGGAGACGATGCTTAACGTTGCGCTT 489
98 ValAspAlaGlyThrValAlaGlyAlaValAsnSerSerValValAs 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 ATAGATGCTGGCAGCTCAGTCAGTCTCTGCA...TCTACACAGCAAA 536
114 pLysSerThrThrPheIleGlyPheSerSerSerSerPheIleAlaSerp 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 TAGAATCTTACTCTCTCAGGTTTCTTACTAGTGTGTTTATCTCTTC 586
131 rGlySerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGly 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
587 CTAGCACAAAGGTTTACTACAGGTGAGGAGACGCTTTC...TCAGCAGA 633
148 SerLeuSerLeuThrLysAsnValSerLeuPheSerLysAsnPhese 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
634 GCGGTAAATTAGAAATAATTCGTAACTGTAGTGTGGGAATTTTC 683
164 rThrAspAsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyT 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
684 TACTCAGATGTGTGAGTATCAAGAGAGCTCTTCTTTAACTGGCA 733
181 hrThmCserAlaLeuPheSerGluAsnThrSerSerLysLysGlyGly 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
734 CTTCGAGAGATGCTCTTTTATGTAACACTCTCATCAACAAGAGAGA 783
198 AlaIleGlnThrSerAspAlaLeuThrIleThrGlyLysGlnGlyGly 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
784 GCAATTGCTTACAGCAGCGCTCGCATAGCAAAATACACACAGGTATGT 833

```

[illegible]

511 ahrtsrlysasnlrleuthlrleuSerGlyThrIleThrleuLeuSprot 528
1698 TGGTTCCAAATAAGATGATCGTATGCTGGGAAGCTCAACCTGATTGATATG 1747
528 hrcglythrphetrlyguasnHissrleuArgaspProGlseryasp 544
1748 AAGGACACTTATTATGAAATCATATGTTCCAGCATGACCAGCTCTTCT 1797
545 lIleuGluLeuLysAlaSerGlyThrValThrSer.....Th 557
574 lnglyThrtrpIyProIleValItrpGlyThrGlyAlaSerThr... 589
1898 AAGACAAATGGAAT...GTATATGTGACATGAGATACGCTCAAAATACA 1944
590AlaThrPheasntrpThrlysrhrglyTYrIleProasnProG 604
1945 AAGAGCGCACGGCACTTGACCAAAACAGGATTGTTCACAGCCCGA 1994
604 uArgIleGlySerLeuValProasnSerLeuTrpAsnaIaPheIleAsp 621
1995 AAGAAATGTGGGTAGTATGCAATACCTTAGGGAGACTTTACTGACA 2044
621 lserSerleuHistrleuMetGluThrAlaAsnGlyLeuGlnGly 637
2045 TTTCGCTCTTGACACAGCTGTGAGATGCGGGCACTGGATGGAACAC 2094
638 AspArgAlaPheTrpCysAlaGlyLeuSerasnPhePheIstLysAsp 654
2095 AAGCAAGTTTCTGGTTTCTTCCTCAATGACAGATCTTCGCAATAGACTGG 2144
654 rThrlysrThrArgGlyPheArgHissrleuSerGlyGlyValIleG 671
2145 AGATGAAATCCCAAGGCTTCGTCATACCTCTGAGAGCTACGATCGAC 2194
671 lGlyIleuLeuHistrpCysSerAspLysIleLeuSerAlaAlaPheCys 687
2195 GGGGAAGTCTCACACTCTTAAGACAGACTATTTACCTTTGCGTTCTGCG 2244
688 GlnLeuPheGlyArgAspArgAspIyPheValAlaLysAsnGlnGlyTh 704
2245 CATCTCTTGTAGAGACAAAGATTGTTTATTCGCTCACAACTACAG 2294
704 rValIyGlyGlyThrLeuTYrGlnHissnGluThr..... 717
2295 AACCTAGCGTGAAGACTTTATTTCTTAACACACTCTATACCTACAAACCC 2344
718TYrIleSerLeu...ProGlyLysLeuArgProCysSerLeuSer 731
2345 AAACATATTGACATTAGAACAGAAAGACTTTTCGATCATGACTATTA 2394
732 TyrValProThrGluIleProValLeuPheSerGlyLysLeuSerTYrTh 748
2395 AAATTCTCCAGGGAATTTCCCTCACCTCTGGATGTCACAAATTTCCCTGAG 2444
748 rHstHrAspAsnAspLeuLYrThrLYrTYrThrTYrProThrVal 765
2445 CCATTGACAAACCGATAGAAACGCACTATATCTTCATTTGCCAAGATCGG 2494
765 yGlySerTrpLysAsnAspSerPheAlaLeuGluPheGlyValArgAla 781
2495 AAGGTCTGTGGAGCAACGAGTATAGCTGGGATGAGCGGCTAGACTT 2544
782 ProIleGlyLeu...AspLysSerAlaLeuPheGlnGlnTYrMetProPh 797
2545 CCTTTGTGTCTTCCAAACCCACACTCTCTTTTTCAGAACCTTACATCA 2594
797 emetLysLeuGlnPheValTYrAlaHissGlnGluIyPheLysGlnG 814

```

2595 GATGAAGTCGAATGGTTTTTATGATCATCAAAATAGCTTCTTCGAAGCT 2644
814  lYthrgluAlaArgGluPheGlySerSerArGleuValAsnLeuAlaLeu 830
2645 CTAAGTAGAGCCGGTGGTTTAAAGTATGGAAAGCGCTAAACCTCGCATY 2694
831  ProIleGlyIleArgPheAspPlyGluSerAspGlySerAlaAspAla...Th 846
2695 CCGTAGGGTGCAGAAATTC...GTCCAGGGGGATATCGAGAGTTCTTACAC 2741
846  rTyAsnLeuThLeuGlyTyThrValAspLeuValArgSerAsnProA 863
2742 CTAATGATCTCTCAGAGATTTCTTTGTTCCGAGTGTATCGTAACATATCCC 2791
863  spcSyrThrThrThrLeuArgIleSerGlyAspSerTrpPlyThrPheGly 879
2792 AATCTTCAGCGAGCATCTTGATGATGAGCCCAACACTCTTGAAATTCGGGT 2841
880  ThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaLeuLysnHisPheCy 896
2842 GGCAACTTTCAGACAGACAGGCATTTTATCTGTAAGGGGTGCAACAACATGCT 2891
896  sPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuArg 913
2892 CTAACAACCTCCAAATGTGAGGCTCTTCGACATCAATCGCTATGAACTCCGTG 2941
913  lYSerSerArgAsnTyrrAsnValAspLeuGlnAlaLysTyrglnPhe 928
2942 GATCTTCAAGAGACTCAATGTAGATGTGGTACCAAACTCCGAATCT 2988

Seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:261508

```

```

seq_documentation_block:
ID      261508 standard; DNA: 3050 BP.
XX
XX      261508;
XX
XX      19-JUN-2000 (first entry)
XX
XX      DNA encoding the CPN100394 polypeptide.
DE
XX      CPN100394; Chlamydia infection; immune response; vaccine; ss
XX
XX      Chlamydia pneumoniae.
OS
XX
XX      Key      Location/Qualifiers
XX      FH      DSD      /tag= a
XX      FT      sig-peptide      101..226
XX      FT      /tag= b
XX
XX      WO200011183-A2.
XX
XX      02-MAR-2000.
XX
XX      18-AUG-1999; 99WO-IB01449.
XX
XX      20-AUG-1998; 98US-0097187.
XX      PR      20-AUG-1998; 98US-0097188.
XX      PR      20-AUG-1998; 98US-0097189.
XX      PR      20-AUG-1998; 98US-0097190.
XX      PR      20-AUG-1998; 98US-0097195.
XX      PR      20-AUG-1998; 98US-0097196.
XX      PR      20-AUG-1998; 98US-0097197.
XX      PR      27-AUG-1998; 98US-0097191.
XX      PR      17-AUG-1999; 99US-0376770.
XX
XX      (CONN-) CONNAUGHT LAB LTD.
XX
XX      Mordin AD, Oomen RP;
XX
XX      WPI: 224703/19.
XX
XX      DR

```

P-PSDB; Y69368.

DR Novel antigens and corresponding DNA molecules that can be used to
XX prevent, treat and diagnose disease caused by Chlamydia infection in
PT mammals, especially humans -
XX
XX
PS
PS

Claim 1; Fig 13A-F; 201pp; English.

CC Z61502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides
CC are present in the bacterial membrane structure, in the external
CC vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane structure
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
XX

S0 Sequence 3050 BP; 897 A; 622 C; 695 G; 836 T; 0 other;

```

alignment_scores:      Quality: 1832.00      Length: 941
                        Ratio: 2.651        Gaps: 23
                        Percent Similarity: 73.433      Percent Identity: 43.571
alignment_block:
US-09-428-122-2 x 261508      ..

```

Align seg 1/1 to: Z61508 from: 1 to: 3050

[illegible]

[illegible]

461 ntfrtgnalaphethrlnglnalalsasrarglucnetalspyalG 478
||||| : : : : : ||| ||| : : : : :
1522 CACACAGAGATTCCTGCAGAGACCAGAAATCCACCGCTCTTTGGATCTGG 15711
478 lythrhrleugluproala...AaplrhserThlIasnasleuval 493
||||| :
1572 GGCCTTCATTAAAGCTTTACAGAGATATTGTATCACAATTTACT 16211
494 lIasnlIeserSerlIeaspylalaIalsylsalaIyslIeglulrThly 510
||||| : : : : : ||| ||| : : : : : ||| : : : : :
1622 ATAAATGCCGATACATTTCACGAAAGAAATCTATCAATATTGTAGCTTC 16711
510 salathrserlysasleuThlleuserglythrIlethrleuasp 527
||||| :
1672 AGCAGGGAATAGAAACATTACCTTAACAGAAAGAACTTACACTTGTAAATG 17211
527 rothrcllythrPheylrGluAsnHisSerleuAlaGAsnProGlnserlyr 543
||||| : : : : : ||| ||| : : : : : ||| : : : : :
1722 CAGATGAGACTTTCTATGAGAACCATACCTTGCAGAGCTCCAGACAT 17711
544 AsplIleuGlulLeuLys...AlaserGlythrValThrSerThrAl 558
:
1772 AGCTTGTAAAGTATCTCCAGAGACGGAGGAGCAATAATTAACACAGA 18211
558 aValThrProAspPro...lIeweclylyuLysPheHisTryGlyT 573
: : : : : ||| ||| : : : : : ||| : : : : :
1822 TCGTTCAGAAAGCCTTCGAAGTAGCTCTTCTAGACACATTATGGCT 18711
573 yrlGlnGlyThrTrpGly...ProIleValTrpGlyThrGlyAlaser... 587
||||| :
1872 ATCAAGGACATTGGATGTGCAAGCATCTCCAGGAACGGAACTCAACGG 19211
588 ThrThrAlaThrPheasnrPrhrlysrThrGlyTrpIleProAsnProGl 604
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1922 AGCAGAGCAAAATTTAGAAAGGGGTGGCGAGAGATACCTTCGAATCCCGA 19711
604 uATrGleGlySerleuValProAsnSerleuTrpAsnAlaPheIleAspI 621
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1972 AGCGGAGAGATCTTAGTTCGCCAATAGCCTGGCGCTCTTTGGTGTATC 20211
621 lIeserSerleuHisTryleuMetGluThrAlaAsnGluGlyLeuGlnGly 637
:
2022 AGCGTCTCTCCAGAAATCAAGGATAGTAATAGAGCCAAATCTTATGTACG 20711
638 AspArgAlaPheTrpCysAlaGlyLeuSerAsnPheHisLysAspSe 654
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
2072 GAAGGGGAGCTCGGGAGCTGAGATTCGTATATTCCTACATAGAGATA 21211
654 rThrlysrThrAlaGlyGlyPheAlaHisLeuSerGlyGlyValIleG 671
:
2122 AATT...AATGAGCAGCGCTATCGCCATAGCGGTCGTTATCTTGGG 21681
671 lyAlaAsnleuHisThrCysSerAspLyslIleLeuSerAlaAlaPheLys 687
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
2169 GAATTGGCCTATGCTTTTGTGATGCTACAGATAAATGGCGCTTTTGGC 22181
688 GluLeuPheGlyAlaGAspArgAspTryrPheValAlaLysAsnGlnGlyTh 704
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
2219 CAGCTCTTCAGTAGAGATTAAGACTACGAGTACCAAAAATCATGTGAC 22681
704 yValTrpGlyGlyThrLeuTrpGlnHisAsnGluThrTrpIleSerL 721
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
2269 TAGCTACTAGGGGCGTATTTCTTGAGAGATACCTTAGATGTTAGAAAT 23171
721 euProLysLysLeuArgProCysSerleuSerTrpValProThGluLe 737
||||| : : : : : ||| ||| : : : : : ||| : : : : :
2318 ..CCACAGGGAATTCACTACTGATACTGCTCAGAAAGCTTGGCTATACCA 23651
738 ProValLeuPheSerGlyAsnLeuSerTrpThrHisThrAspAsnAspLe 754
||||| :
2366 GTCGCTACACTAGATATGAGATGCTTACAGCCCTAAGAAATATGATAT 24151
754 uLysTrIleValTrpThrThrTrpProThValGlySerTrpGlyAsnA 771

```

2416 GAAACCAATACACGACATATCCAGAAAGCTCAGGATCTGGGCAATG 2465
771 spSerPheAlaLeuGluPheGlyAlaValPheIleCysLeuAspGlu 787
2466 ATGTTTGGTCTTGAGTTGGACGACTACTACTACTACCTTAACAGT 2515
788 SerAlaLeuPheGluGlnTyrMetProPheMetLysLeuGlnPheVal 804
2516 ACTTTTATTTATTTACTCTCCGTTTCTTCAGGCTGAGTGCACCTA 2565
804 ValAlaGlnGlnGluPheLysGlnGlnGlyThrAlaAlaArgGluPhe 821
2566 TGCCTACAGAGAACTTCAAGAGACAGAGAGTGCAGTTCGTACTT 2615
821 LysSerArgLeuValAlaSerLeuAlaLeuProIleGlyIleArgPhe 837
2616 CTGCGGAGATCTTTTCAATTTACAGTTCTCTATTTGGCTGAAGTT 2665
838 LysGluSerAspCysGlnAspAlaThrTyrAsnLeuThrLeuGlyTyr 854
2666 AGATTTTACAGACTGTAAAGGGATCTTATGACTTACCTTGCTATGT 2715
854 ValAspLeuValArgSerAsnProAspCysThrThrLeuArgLys 871
2716 TCCTGATGATTCGCAAGATCCCAAGACAGCGCAACATTTG...GCTA 2762
871 ArgLysAspSerTyrLysThrPheGlyThrAsnLeuAlaArgGlnAla 887
2763 GTGAGAGTACGTGAGACACCCAGCAACATCTCTCCACACAGGATTA 2812
888 ValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsnPheGlnAla 904
2813 CACTGGGTTAGGAACCACTGTCTCATTAATCTCGAATTAGGTGT 2862
904 eSerGlnPheSerPheGluLeuArgLysSerArgAsnTyrAsnVal 921
2863 CAATTCACGGAGTATTGAATTCGGGGATCTCTCGTAATTATACATCA 2912
921 splLeuGlyAlaLysTyrGlnPhe 928
2913 ATCTCGGGGTAAATACCGATTT 2935

```

seq_name: /SIDS6/gcdata/geneseq/geneseq/NA1999.DAT:X06827

seq_documentation_block:

ID X06827 standard; DNA; 2838 BP.

AC X06827;

26-APR-1999 (first entry)

Chlamydia pneumoniae surface exposed protein Omp15 DNA.

Omp15: outer membrane protein 15: surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

Chlamydia pneumoniae.

W05858953-A2.

30-DEC-1998.

19-JUN-1998; 98WO-DK00266.

23-JUN-1997; 97DK-0000744.

(BIRK/) BIRKJUND S.
(CHR/) CHRISTIANSEN G.

Birkjund S, Christiansen G, Knudsen K, Madsen A;

Mygind P;

PI

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

WPI: 1999-105610/09.
P-PSDB: W88428.

Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins

Claim 6; Page 70-71; 115pp; English.

This DNA sequence codes for the novel surface exposed protein Omp15 (see W88428) of the human respiratory pathogen Chlamydia pneumoniae. By generating antibodies against C. pneumoniae outer membrane complex, a polyclonal antibody (PAB 150) was obtained which reacted with outer membrane proteins. The antibody was used to identify the genes (see X06816-27) encoding Omp4-Omp5 proteins (see W88417-28) in an expression library of Chlamydia pneumoniae DNA. The genes are situated in 2 gene clusters: Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in the other, and encode polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The invention provides a new species specific test for identifying mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp5 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.

Sequence 2838 BP; 837 A; 591 C; 648 G; 762 T; 0 other;

alignment_scores: Quality: 1811.00 Length: 945

Ratio: 2.621 Gaps: 24

Percent Similarity: 73.122 Percent Identity: 42.963

alignment_block:

US-09-428-122-2 x X06827 ..

Align seg 1/1 to: X06827 from: 1 to: 2838

```

22 AlathrGluThrValLeuAspSerAlaSerPheAspGlyAsnLysAs 38
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 GCCGCACACACTCCACATAATCTGAAGATGGTTTATGGGAGGCGCAA 116
38 nGlyAsn...PheSerValArgGluSerGlnGluAspAlaGlyThr 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 TACAATATCTTTTCTCCGAATCTACACAGGATGTCGAGAGCTACCT 166
54 YrLeuPheLysGlyAsnValThrLeuGlnAsnIleProGlyThr 70
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 ACTCTCTACAGAGAGGTCTGTATTAGAT...CCGGGAAAGGTGT 213
71 AlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThrPhe 87
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 TCAATTACAGGAAGTCTGTGTAGAAAGTGTGCGCATTTTCATTTT 263
87 rGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAlaGly...Thr 103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 AGTAAATGAAATACCTTAAGTTCTGTGAGATGACGAGGTCTATATA 313
103 aLaIleGlyAlaIleValAsnSerSerValValAspLysSerThrPhe 119
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TCCGGTGTCTCATGTACAGAAAGT.....AAGATTTAAGTTC 354
120 IleGlyPheSerLeuSerPheIleAlaSerProGlySerSerIleTh 136
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ACAGATTCTCTTCTCTGTGATCACAGAAATCTCAAAATCCGCTGTAG 404

```



```

2157 AATTCATGAGCACTACTACTCAGGGGCTGTAATTTCTTGAGGATACCCCTAG 2206
717 hrTyrlleSerLeuProCysLysLeuArgProCysSerLeuSerTyVal 733
2207 AGTTTAGAAGT...CCACAGGGAATCTTACTGATGATCTCCTCAGAGCT 2253
734 ProThGluLeProValLeuPheSerGlyAsnLeuSerTyThrHisH 750
2254 TGCTGTACCAAGCTGCTACATATGATATGCAAGTGTGCTTACAGGCATAG 2303
750 rAspAsnAspLeuLysThrLysTyThrThrThrTyProThValLysGly 767
2304 AATATATGATGATGAAACCAATACAGACATATCCAGAAGCTCAGAGAT 2353
767 exTrpGlyAsnAspSerPheAlaLeuGluPheGlyGlyValArgAlaProIle 783
2354 CTTCGGCAATATGATGTTTTCGCTTATAGTTTGAGGACGACTACATATAC 2403
784 CysLeuAspGluSerAlaLeuPheGluGlnTyMetProPheMetLysIle 800
2404 TACCTTACACTACTCTTTTATTTGATGATACTACTCTCCGTTCTCAGCT 2453
800 uGlnPheValTyraIaHisGlnGlnGlyPheLysGlnGlnGlyThrGluA 817
2454 GCAGTGCACCTATGCTCACCAGGAAAGCTTCAAAAGAGACAGGAGCTGAGG 2503
817 lAaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGly 833
2504 TTCGTCACTTACTAGGCGGAGATCTTTCAATTGACATTCCTCTTCTGCG 2553
834 lIeaArgPheAspLysGlnSerAspCysGlnAspAlaThrTyAsnLeuTh 850
2554 GTGAGATTTGAGAGATTTTCAGACCTGAAGGAGATCTTATGAACTTAC 2603
850 rLeuGlyTyThrValAspLeuValArgSerAsnProAspCysThrThrt 867
2604 CCTTCCTTATGTTCTCCGATGTGATGTGCAAAATGATCCCAAGACGCGCAA 2653
867 hrLeuArgIleSerGlyAspSerTyThrLysThrPheGlyTyThrAsnLeuAla 883
2654 CATTG...GCTACTGACACTACGTGAGAGACCCACGAAACAATCTCTCC 2700
884 ArgGluAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAs 900
2701 AGACAAAGATATACACTGCGTTTACGAGAACACACTGTCATATAATCCTGG 2750
900 nPheGluAlaIleSerGlnPheSerPheGluLeuArgGlySerSerArg 917
2751 AATTGAGGTTCACAGCACGAGCATATGATTCGGGGATCCCTCTCGTA 2800
917 snTyAsnValAspLeuGlyAlaLysTyArgLysPhe 928
2801 ATTATATACATCAATCTCGGGGTAAATACCGATT 2835

seq_name: /SID56/gcdata/geneseq/geneseqn/NA199.DAT:X06819
seq_documentation_block:
ID X06819 standard; DNA; 2526 BP.
XX
XX X06819;
XX
XX
XX 26-APR-1999 (first entry)
XX
XX Chlamydia pneumoniae surface exposed protein Omp7 DNA.
DE
XX Omp7; outer membrane protein 7; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX Chlamydia pneumoniae.
OS
XX
XX W09858953-A2.
XX
XX 30-DEC-1998.
XX
XX
XX

```

```

XX PF 19-JUN-1998; 98MO-DK00266.
XX PF
XX PR 23-JUN-1997; 97DK-0000744.
XX PR
XX PA (BIRK/) BIRKELUND S.
XX PA (CHRL/) CHRISTIANSEN G.
XX PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX PI Mygind P;
XX DR MPI; 1999-105610/09.
XX DR P-PSDB; W88420.
XX
XX PT Species-specific test for identifying mammals infected with
XX PT Chlamydia pneumoniae - comprises detecting antibodies specific for
XX PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX PT these proteins
XX PS
XX PS Claim 6; Page 49-50; 115pp; English.
XX
XX This DNA sequence codes for the novel 89.7 kDa surface exposed
XX protein OmpF (see W88420) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06816-27) encoding
XX Omp4-Omp5 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;
XX
XX
XX Alignment_scores:
XX Quality: 1564.00 Length: 942
XX Ratio: 2.523 Gaps: 20
XX Percent Similarity: 65.817 Percent Identity: 39.172
XX
XX Alignment_block:
XX US-09-428-122-2 x X06819 ..
XX
XX Align seg 1/1 to: X06819 from: 1 to: 2526
XX
XX 12 ThPhAlaIlePheProLeuSerMetIleAlaThrGluThrValLeuAs 28
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 43 ACCCTTCTATCTCCATTATTAGAGCGTCTACTACCGAAGGACTATC 92
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 28 pSerSerAlaSerPheaspGly..AsnYasngIlyasnpheSerVala 44
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 93 GCGTACCAATAGCTTCGATGGAGACTCATCACAACAGCTTTCTTCTA 142
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 44 rglUserGlnGluaspAlaGlyThThTyrLeuPheIyAsnVal 60
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 143 AAACATCATTCGGCTACAGATGCGACCAATATGTTTAAAGATTCGTA 192
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 61 ThleuGluAsnIlePheogIlyThrAlaIleThrIySerocysh 77
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 193 GTTATAGAAATCTACCCAAACAGGGAAGTCAAGTCTACTAGTTGTT 242
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 77 easnaSn.....ThIySgIyAspleuThrPheThnGlyAsnGlyAsnS 92

```

243 TAAATGACCGCTGACCTGAGATCTAAATTTTGTAGGAGGGGATTTT 292
92 erleuLeupheglnthrvAlaSpAlaGlythrvAlaGlyAlaAlaVal 108
CTTACATTTAGCAATATCGATCGAACACGCGCTTGGAGCTGCTAT 342
109 AsnSerSerValAlaAspLysSerThrThrPheIleGlyPheSerSer 125
343 GGAAGTACACGACTAATAGACAGTCAGCTTATCAGCATTTTGGCAGT 392
125 userPheIleAlaSerProGlySerSerIleThrThrGlyGlyAlaVal 142
393 TTCTTTCTTAATATCCCGACAGATACAGTACATGATGATGGAGACT 442
142 aLserCysSerThrGlySerLeuSerLeuThrLysAsnValSerLeu 158
443 TCATGCTTAA...GGGAATTTAAGCTTATGATATGATAGATATG 489
159 PheSerLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysTh 175
490 ATTGAGACATTTCTCAACAGAGATGGGAGCAATTAATGT... 534
175 rLeuSerLeuThrGlyThrMetSerAlaLeuPheSerGluAsnThr 192
534 534
192 erSerLysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIle 208
534 534
209 GlyAsnGlnGlyValSerPheSerAspAsnThrSerSerAspSerG 225
534 534
225 yAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAlaLys 242
535GCAGGCTCTTGAAGATCGCAACAAATAGTCC 568
242 aLserPheIleAspAsnLysValThrGlyAlaSerSerSerThrThr 258
569 TTTCTTTTAT... 579
259 AspMetSerGlyAlaIleCysAlaLysThrSerThrSerThrLys 275
579 579
275 sValThrLeuThrGlyAsnGlnMetLeuPheSerAsnAsnThrSer 292
580GCAATAGTCTT 592
292 hrThrAlaGlyGlyAlaIleTyrValLysLysLeuGluLeuAlaSerG 308
593 CAACACGTGGCGGAGCATTCATACCAAAACCTCACACATATCTTGT 642
309 GlyLeuThrLeuPheSerArgAsnSerValAsnGlyLysThrAlaPro 324
643 GGGGAACCTCATTT...CAGGGGAATACAGCGCTTAC 677
325LysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeu 338
678 GGCTGCTGTTAAAGAGGTCTATCGGATTGCGACTCTGGCCTCTAT 727
338 erLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 354
728 CCATTTCTGAGACAGTGGCAGCATTTATCTTGAAGCATACATAGCA 777
355 SerThrThrProGlyThr...AsnArgSerSerIleAspLeuGlyThrSe 370
778 GCTACA...GGAACTGCTCTCATATGCTATGATTTAGGAACCTAG 821
370 rAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPhe 387
822 CGCTAAGATACGCTGTTACGCTCTGCGCAGACATACATATCTTT 871
387 yAspProIleThrThrGlySerSerThrThrValThrAspValLeuLys 403
872 ATGATCGATTTACTGTATACAGATCGACATCTGTCTGATGCTCAAT 921
404 ValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIle 420
922 ATTAATAAGCCCTGATCTGAGATTAACAAAGATATACGGGAACCAT 971
420 ePheThrGlyGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsn 437
972 CTTTCTCGGAGACAGCTCCGAGGACAGACTAAAGATGAAGAAGACC 1021
437 eutThrSerLysLeuGluProValThrLeuSerGlyThrLeuSer 453
1022 GCATCTTAATATTAATCAAAATGTTGCTTTAAATAGGAGCATGAT 1071
454 LeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrGlnGlnAla 470
1072 TTAAGAGTATGCTGTTTAAGTGGACAGCGTTTCTCAGATGCAAA 1121
470 pSerArgLeuGluMetAspValGlyThrThrLeuGluProAlaAspThr 487
1122 CTCTAAGTATGATGATGATTTAGGAGCTGTTG...GTTGCAACACCG 1168
487 erThrIle...AsnAsnLeuValIleAsnIleSerSerIleAspGly 501
1169 AAGTATCGAGTTAAGCAATTTGAAATTAATATACCTCTCAGAGAC 1218
502 AlAlaLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThr 518
1219 GGGAAAAAGATTAAGTCACTGCTGCCACAGCTCAGAAAGATTTGCTAT 1268
518 uSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsn 535
1269 AGATCGTCTGTTGCTGCAATGAGATGAGATTTTATCAAAAG 1318
535 lAserLeuArgAsnProGlnSerTyrAsp...IleLeuGluLeuLysAla 550
1319 GCTTTTGAATGAGACCATCTCTATGATGGAGTCTTGATGATGCT 1368
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
1369 GGAAGACATCGTGAATTTCTGACATCTCGCAGT...ATAATGCTGT 1415
567 uLysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGly 584
1416 ACAATCTCCGTATGCTATCAGGGAAGTGG...ACAATCAATGGTCTA 1462
584 hrGlyAlaSerThrThrAlaThrPheAsnThrPheLysThrGlyTyrIle 600
1463 CTAT...GATTAAGAAAGCTACGCTTTCTTGCGCAAGCAAAATTTAAT 1509
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
1510 CCCACTGCTGACGAGGAGGCTCGTATGCTTAATCTTTGGGGTTC 1559
617 aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGlu 634
1560 TTTTATGATGTTGTCCTCCCAAAATTTATAGAGCTAGTACGAAG 1609
634 LysGlnGlnLysArgAlaPheThrCysAlaGlyLeuSerAsnPhePhe 650
1610 GTCTCTCTTAAGAAAGATTTTGGTGGTGAAGCATTTCCATGTTTG 1659
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGly 667
1660 CATAGAGCGGTGTGTAATAATCAAGAAATTCGTCATGTAGTGAAG 1709
667 yTyrValIleGlyLysLeuHisThrCysSerAspLysIleLeuSer 684
1710 TGCTGTAGTAGTGTACGACGAGATGCCGGGTGTGATCTTGTCTC 1759

```

684 laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
      ::::::::::::::::::::::::::::::::::::::::::::::
1760 TGGATTTCCTCACTCTTTGGCGGTGACAAAGACTACTATTATGAATACC 1809
      ::::::::::::::::::::::::::::::::::::::::::::::
701 AengInGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGluTh 717
      ::::::::::::::::::::::::::::::::::::::::::::::
1810 AATTTCGCAAAAGCCTACGACAGATCTTACGTTGACGACGATGCTTC 1859
      ::::::::::::::::::::::::::::::::::::::::::::::
717 TTYrILeSerLeuProCys.....LysLeuAlaGProC 728
      ::::::::::::::::::::::::::::::::::::::::::::::
1860 CCTATACCTCTGTGAGTATCCTTTAGAGAGAGGAGGACTCGCGAGA 1909
      ::::::::::::::::::::::::::::::::::::::::::::::
728 ySSerLeuSerTyrValProThrGluLeuProValLeuPheSerGlyAsn 744
      ::::::::::::::::::::::::::::::::::::::::::::::
1910 TCCTGTTCCTTATGTTCCAAGACTCTCCGCTGCTCTTCTATAGCGAG 1959
      ::::::::::::::::::::::::::::::::::::::::::::::
745 LeuSerTyrThrHisThrAspAsnAspLeuLysThrLysTyr..... 758
      ::::::::::::::::::::::::::::::::::::::::::::::
1960 CTTAGCTAGCGCCATACGATACGATCGATGAAGACCGAGTCTTACCCC 2009
      ::::::::::::::::::::::::::::::::::::::::::::::
759 .....ThrTyrProThrValLysGlySerTyrGlyAsnAspSerP 773
      ::::::::::::::::::::::::::::::::::::::::::::::
2010 CCCCCCCCCGCGCTCTCGAGGATCATCTTTGGGAGGATATGCT 2059
      ::::::::::::::::::::::::::::::::::::::::::::::
773 heAlaLeuGlnPheGlyGlyArgAlaProIleCysLeuAspGluSerAla 789
      ::::::::::::::::::::::::::::::::::::::::::::::
2060 GGGCTGAGAGCTGGGAATCGA.....GTGCTGTGAAAATACACAGC 2103
      ::::::::::::::::::::::::::::::::::::::::::::::
790 .....LeuPheGlnGlnTyrMetProPheMetLysLeuGlnPheVa 803
      ::::::::::::::::::::::::::::::::::::::::::::::
2104 GGCAGAGATTTTCGAGACTACATCTTGTAAAGCCAAAGCTGT 2153
      ::::::::::::::::::::::::::::::::::::::::::::::
803 lTYrAlHisGlnGlnGlyPheLysGlnGlnGlyThrGluAlaArgGluP 820
      ::::::::::::::::::::::::::::::::::::::::::::::
2154 TTACTCGCGCAAGATAGCTTTGTTGAACAGAGAGCATACAGTCGAT 2203
      ::::::::::::::::::::::::::::::::::::::::::::::
820 heGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyIleArgPhe 836
      ::::::::::::::::::::::::::::::::::::::::::::::
2204 TTAGTGTTCGACATCTTATTAACCTTCGATCTCTTGAATCAAGTTA 2253
      ::::::::::::::::::::::::::::::::::::::::::::::
837 AspLysGlnSerAspCysGlnAspAlaThrTyrAsnLeuThrLeuGlyTyr 853
      ::::::::::::::::::::::::::::::::::::::::::::::
2254 GAGAAACGG.....TTTGCAGAGCATATTAATCATGTTTACGATGTA 2297
      ::::::::::::::::::::::::::::::::::::::::::::::
853 rThrValAspLeuValArgSerAsnProAspCysrThrThrLeuArgI 870
      ::::::::::::::::::::::::::::::::::::::::::::::
2298 TTCTCCAGATGTTGTCTAGTAACCCAAATGTACGACTACCTTACTT 2347
      ::::::::::::::::::::::::::::::::::::::::::::::
870 leSerGlyAspSerTyrPlysThrPheGlyThrAsnLeuAlaArgGlnAla 886
      ::::::::::::::::::::::::::::::::::::::::::::::
2348 CCAACCAAGGAGCTTGAGAGACCAAGGTTGCAACTTACCAACAGCGCT 2397
      ::::::::::::::::::::::::::::::::::::::::::::::
887 LeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsnPheGluAl 903
      ::::::::::::::::::::::::::::::::::::::::::::::
2398 GGTATTGTCAGGCGCTCAGGTTTTCGATCTTTGGAGCTGCACAGAGCT 2447
      ::::::::::::::::::::::::::::::::::::::::::::::
903 apHeSerGlnPheSerPheGlnLeuArgGlySerSerArgAsnTyrAsnV 920
      ::::::::::::::::::::::::::::::::::::::::::::::
2448 TTTTCGGGAACTTTGGCTTGAATGGCGGAGATCTCTCTACTATAAG 2497
      ::::::::::::::::::::::::::::::::::::::::::::::
920 alAspLeuGlyAlaLysTyrGlnPhe 928
      ::::::::::::::::::::::::::::::::::::::::::::::
2498 TAGATCGGGTAGCAAAATCAATTTT 2523
      ::::::::::::::::::::::::::::::::::::::::::::::

```

```

seq_name: /SIDS6/gcgdata/geneseq/NA1999.DAT:X06818
seq_documentation_block:
ID X06818 standard; DNA; 3052 BP.
XX
AC
XX X06818;
DT 26-APR-1999 (first entry)

```

```

XX XX Chlamydia pneumoniae surface exposed protein Omp6 DNA.
DE
XX
XX Omp6: outer membrane protein 6; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX Chlamydia pneumoniae.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..2764 /*tag= a
XX
XX MO9858953-A2.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-DK00266.
XX
XX 23-JUN-1997; 97DK-0000744.
XX
XX (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A,
XX Mylind P;
XX WPI: 1999-105610/09.
XX P-PSDB: W88419.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 45-46; 115pp; English.
XX
XX This DNA sequence codes for the novel 100.3 kDa surface exposed
XX protein Omp6 (see W88419) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06818-27) encoding
XX Omp4-Omp15 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
XX the other, and encode polypeptides of about 89,6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 3052 BP; 875 A; 603 C; 653 G; 921 T; 0 other;

```

```

alignment_scores:
Quality: 1430.50 Length: 951
Ratio: 2.187 Gaps: 26
Percent Similarity: 68.770 Percent Identity: 36.698

```

alignment_block:

US-09-428-122-2 x X06818 ..

Align seg 1/1 to: X06818 from: 1 to: 3052

```

5 PhePro..LysPheValPheSerThrPheAlaIlePheProLeuSerIle 20
||||| ||| :::::::::::::: ||:::

```

```

22 TTCTCTACTTTTCTTTAACAATGCTCTGAGTCTTGACACTTCTT 71
20 ttleAlaThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnL 37
   : ||||| ||| :||| ||||| |||||
72 GAGTGTACTACGATTTCTTTAACCCAGAAAGATAGTTTCATGCGAGATA 121
37 ysaSnGlyAsnPheSerValArgGluSerGlnGluAspAlaGlyThrThr 53
   :||| :||| :||| :||| :||| :|||
122 GTGAGAAAT.....GCAGAACGTTCTATATATGTTCAAGCTGGGAGATGTC 165
54 TyrLeuPheGlyGlyAsnValThrLeuGluAsnIleProGlyThrGlyTh 70
   ||| :||| :||| :||| :||| :|||
166 TATAGCCTTACTGTGATGTCTCAATATCTAACGTC.....CATAGTC 209
70 rAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThrPhe 87
   :||| :||| :||| :||| :||| :|||
210 TGCATTAAATAAAGCCTTGCTTCAATGTGACCTCAGAAAGTGTGACGTTG 259
87 hrcLysnGlyAsnSerLeuLeuPheGlnThrValAspAlaGlyThrVal 103
   :||| :||| :||| :||| :||| :|||
260 CAGGAATCATCATGGGTATATTATTAATAATTCTCAGGAACATACA 309
104 AlaGlyAlaValAsnSerSerValAlaAspLysSerThrThr...Ph 119
   :||| :||| :||| :||| :||| :|||
310 AAGGAAGGCTGTACTTCTTGCCAAAGATCCTCAACACGCGACGCTT 359
119 eileGlyPheSerSerLeuSerPheIleAlaSerProGlySerIle 136
   :||| :||| :||| :||| :||| :|||
360 TTCTGGTCTCCACGCTCTCTTTATTCAGAGCCCGGAGATATTAAAG 409
136 hrrhGlyLysGlyAlaValSerCys....SerThrGlySerLeuSer 150
   ||| ||| ||| ||| ||| |||
410 AACAGGA.....TGCTCTATTCAAAAAAATGCACATTATG 444
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 167
   ||| :||| :||| :||| :||| :|||
445 CTCTTAACATATATAGTAGCGTTTGAACAAACCAAGTAAGACTAA 494
167 nglGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMet 184
   :||| :||| :||| :||| :||| :|||
495 AGCGGCGCATTAAGTGGGCGAATGTACTATAGTAGCACTACGATT 544
184 eAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
   ||| :||| :||| :||| :||| :|||
545 CCGTCTCTTCTATCAAAATGCAGCCACT...TTTGAGGTGCTATTCAT 591
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPhe 217
   :||| :||| :||| :||| :||| :|||
592 TCTTCAGATCCCTACAGATTGCAGTAATCAGACAGATAAGATTGC 641
217 rAspAsnThrSerSerAsp...SerGlyAlaAlaIlePheThrGluAla 233
   :||| :||| :||| :||| :||| :|||
642 ACAAAATACGCCAGAAATGTTCTGGAGGGCTTGTACTCCGATGCTG 691
233 eAlaThrLysSerAsnAsnAlaLysValSerPheIleAspAsnVal 249
   :||| :||| :||| :||| :||| :|||
692 ATATTGATATGATCAGAAATGCTTATCTATTTCAGAAATGAG... 738
250 ThrGlyAlaSerSerThrThrGlyAspMetSerGlyAlaIleCys 266
   :||| :||| :||| :||| :||| :|||
739 .....GCATTGACTACTGCTATAGTAAGAGAGGGCTGCTG 776
266 sAlaTyrLysThrSerThrAspThrLys.....ValThrLeu 279
   :||| :||| :||| :||| :||| :|||
777 TTGTCTTCCACTTCAGAGAGTAGTACTCCAGTTCCTATTGTGACTTCT 826
279 hrcLysnGlnMetLeuLeuPheSerAsnAsnThrSerThrThrAlaGly 295
   :||| :||| :||| :||| :||| :|||
827 CTGACATAAACAAGTATAGTCTTGAAGAAACCATTCATTAATGGGCGC 876
296 GlyAlaIleTyrValLysLysLeuGlnLeuAlaSerGlyLysLeuThr 312
   :||| :||| :||| :||| :||| :|||
877 GGAGCATTTATGCTAGAAACTTAGCATCTTTCAGAGAGTCTACTCT 926

```

```

312 upheSerArgAsnSerValAsnGlyThrAlaProLysGlyGlyAlaI 329
   :||| :||| :||| :||| :||| :|||
927 ATTATCATTAATATATATCATATGCAAAATTCGCAAAATTAGGTGGAGCTA 976
329 leAlaIleGluAspSerGlyGluLeuSerLeuSerAlaAspSerGlyAsp 345
   :||| :||| :||| :||| :||| :|||
977 TTGCCATTATACTGAGGGGAGATCATGTTATTCACAGAGAAAGAGACA 1026
346 lIeValPheLeuGlyAsnThrValThrSerThrProGlyThrAsnAr 362
   :||| :||| :||| :||| :||| :|||
1027 ATTACATTCCAGAAAC.....CGACGAGCTTACCGTTTTCAT... 1068
362 gSerSerIleAspLeuGlyThrSerAlaLysMetThrAlaLeuArgSer 379
   :||| :||| :||| :||| :||| :|||
1069 ....GCAATCCACTTCTTACAAAATGCTAAATTCGTAATTACAGCGCA 1114
379 lAlaLysArgAlaIleTyrPheTyrAspProIleThrThrGlySerSer 395
   :||| :||| :||| :||| :||| :|||
1115 GAATGGAATGCTTATAGAATTTTATGATCTATTAAT...TCTGAA 1158
396 ThrThrValThrAspValLeuLysValAsnGluThrProAlaAspSer 412
   :||| :||| :||| :||| :||| :|||
1159 GCAGATGGGTCTACCAATTCATATCAACGAGAGATCCTAAATAATAA... 1206
412 aLeuGlnTyrThrGlyAsnIleIlePheThrGlyGlyLysSerGly 429
   :||| :||| :||| :||| :||| :|||
1207 ...GAGTACACAGGAGCATTCTTCTGAGAAAG.....A 1243
429 hrcGluAlaAspSerLysAsnLeuThrSerLysLeuLeuGlnProVal 445
   :||| :||| :||| :||| :||| :|||
1244 GTCTACAAACGATCCTAGGATTTAAATCTAACATCCCTCAGAACGTC 1293
446 ThrLeuSerGlyGlyThrLeuSerLeuLysHisGlyValThrLeuGln 462
   :||| :||| :||| :||| :||| :|||
1294 AACCTGTGCAAGATATCTAGTATTAAAGAGGGCGGAGATCACT 1343
462 rGlnAlaPheThrGlnGlnAlaAspSerArgLeuGlnMetAspValGly 479
   :||| :||| :||| :||| :||| :|||
1344 TTCAAATTCACGACATCTCCAGAGATGCATTTAGTTAGATTGAGTA 1393
479 hrrhLeu...GluProAlaAspThrSerThrIleAsnAsnLeuVal 494
   ||| ||| ||| ||| ||| |||
1394 CCAAACTGATAGCCTTAAGAAAGACATTGCCATCAGCGCTCGGATA 1443
495 AsnIleSerSerIleAspGlyAlaLysLysAlaLys...IleGluThr 510
   :||| :||| :||| :||| :||| :|||
1444 GATATGATAGCTTAGCTATCATCTCAACAGCAGCTGTTATTAACAAA 1493
510 sAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeuAsp 527
   :||| :||| :||| :||| :||| :|||
1494 CACCGCAAAATAACAGATATCCGTGAGGAGCTTATAGAACTTATCTGC 1543
527 rThrGlyThrPheTyrGluAsnHisSerLeuArgAsnProGlnSer 543
   :||| :||| :||| :||| :||| :|||
1544 CTACTGGCAATGCCATGAAGATCTCAGAAATGGAATTCACAGACGTC 1593
544 AspIleLeuGluLeuLys.....AlaSerGlyThrValThrSerThr 558
   :||| :||| :||| :||| :||| :|||
1594 CCTCTGCTCTTATAGAGCTGAGCGGGGAGTACGTAAGTACTGC 1643
558 a.....ValThrProAspProIleMetGlyGlyLysPhe 570
   :||| :||| :||| :||| :||| :|||
1644 TGGAGATTTCCTACCGGTAGTCC.....C 1669
570 lstrGlyTyrGlnGlyThrThrGlyProIleValTyr...GlyThrGly 585
   :||| :||| :||| :||| :||| :|||
1670 ATATGCTTTTCAAGCAATGG...AAATAGCTTGAGCAGAGAACTGGA 1716
586 AlaSerThrAlaThrPheAsnTyrThrLysThrGlyTyrIleProAs 602
   :||| :||| :||| :||| :||| :|||
1717 ...ACCAAAAGTGGAAATTCCTGCGATAAAATTAATTAAGCTAG 1763

```


Align seg 1/1 to: A27342 from: 1 to: 3150

```

1 MetLysSerSerPheProLysPheValPheSerThr..... 12
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ATGAAACGCTATTCGTAGTTCTTAATTCTACACACTGGCGCATG 150
13 PheAlaIlePheProLeuSerMetIleAlaThrGluThrValLeuAspS 29
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 TTTTCTCT.....TCACACACCTTTTACTGTAGAGTATCATCTC 188
29 eSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheserValArgLys 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 CTCCGAGACACTTGATGATCGAGGAGGAGATTTTCTCTACACAA 238
46 SerGlnGluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 CTTCGTGATCTAGAGGACACTCTGTATTTTTCAGGGATCTCTACAT 288
62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnA 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 TCCGAACTCTGATTAATGCGCATATCCAGAACCTCTCCAGTTCCTTA 338
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuPhe 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ATAGGGCGGAGACACTACAATCTTAGAAGAGTGGGTTTCTCTC 388
96 GlnThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerVa 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 TTTAAATATC...CGTTCTTCAGCTGACGGAGCCGCGATTAGTAGTAT 435
112 LValAspLys.....SerThrPheIleGlyPheSerSerL 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 CACCCAAAATCTCTGAACATATGTCCTTGAGTTTTCAGGATTTAGTCA 485
125 euSerPhe.....IleAlaSerProLysSerSer 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 TGATCTTCGATACGTGTAATCTTGACTTCAGATACCTCAGCGAGTAT 535
135 IleThrThrGlyLysGlyAlaValAsnSerSerThrGlySerLeuSer 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 GTCATACCTCAGCGCATCGGCGATTTACGCTACACGCCCATGCT...TT 582
151 uThrLysAsnValSerLeuLeuPheSerLysAsnPheserThrAspArg 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 TACAACAATGACTCCATCTCATCTCAATACACCGCTTCTCGACGATTG 632
168 LysIleAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSer 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 GAGCTGCATTCGAGGACACACATCACAATAGAAAATACGAAAAGAC 682
185 AlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGlnThr 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 CTTCCTTTAAATGTAATGATGATCTCTAATGAGGGGCCCTCAGCGG 732
201 rSerAspAlaLeuThrIleThrGlyAsnGlnGlyLysValSerPheSer 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733 ATCTCAGGATCAACCTCATCAACATAGCGCTCTGATTTCTCA 782
218 sp..... 218
783 CGAATGTACAGAGATCTATGTGGGCTATTACCTACCGAGAGATCT 832
219 .....AsnThrSerSer 222
833 ATGCTACCTCTGGAGACCTCTCAGAGAGTCTTGTCGTTAATATAGCTC 882
222 rAspSerGlyAlaAlaIlePheThrGluLysSerValThrIleSerAsn 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
883 GCGCTCAGAGGCGCTATCTATGTAACGAAATGTCACATTTTCTAA 932
239 snAlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSer... 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

933 ACAGCGACCTGACTTCCAAACAATACAGCATCTCCCAAAACTCCTTA 982
255 .....SerThrThrGlyAspMetSer 261
983 CCTGCACCTACACTCCACCTACACACAGAGAGTCACTCTTGTAGG 1032
261 r...GlyGlyAlaIle...Cys.....AlaTyrLysThrSerThr 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1033 ATATGAGCGCCCATCTTCTACTCTCCAGCTACCCGCCACACACAG 1082
273 spThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsn 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1083 GTGTTAGCTGACTATATCTGGAGAAAACAGCTTACTTCTTAGAAAC 1132
290 ThrSerThrThrAlaGlyAlaIleTyrValLysLysLeuGluLeu 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1133 ATTGCTCCGAAACAGAGAGAGCCCTCATGCAAAAAGATCTATAGA 1182
306 aSerGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1183 TTTATATTAATCTACAAATATTTCTT.....GGAATACAG 1217
323 LaProLysGlyAlaAlaIleAlaIleGluAspSerGlyLysLeuSerLeu 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 CTGAAAGAGAGCGCCCTATGCTATCCGAATCTGGGAGCTCTCTCTA 1267
340 SerAlaAspSerGlyAspIleValPheLeuGlyAsn...ThrValThrSe 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 TCCGCAATCAAGGATATCTCTTTAACAAAGAACCTCACATCCTAC 1317
355 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 TGGGACACT....ACTCCGAATAGATTTCACTTCGGAAGAAAGTCCA 1361
372 ysmetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheTyrAsp 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1362 AGTTGCCACTCTAGAGAAATCGCAGAGCTATACCTTACTTCTATGAT 1411
389 ProIleThrThrGly.....SerSerThrThrValThrAspValLeu 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1412 CCGATTACATGATGATTTATCTGCTCATCCGAGCGCTACTGCT 1461
403 sValAsn...GluThrProAlaAspSerAlaLeuGlnIleThrGlyAsn 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1462 COTCAATCCCAAGCAGCTGAGATGCTCG...TATTCAGGAGCTA 1505
419 IleLeuPheThrGlyLysLeuSerGluThrGluAlaAlaAspSerLys 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1506 TTGCTTTTTCAGAGAAAACCTCTCACTGACGAAAGCAACCCCTCA 1555
436 AsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLe 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1556 AATGCTACATCTACATTAACCAAGAGCTGAAGCGGAGTCT 1605
452 uSerLeuLysHisGlyValThrLeuGlnThrGluAlaPheThrGlnGln 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1606 CGCTTAAGAAGCGTGTACTTAAATGTTCAATCACTTCCAGCAAGT 1655
469 laAspSerArgLeuGlnMetAspValGlyThrThrLeuGluProAlaAsp 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1656 AAAAGTCCGTCATCATGATGACGAGGACACATTTGCAACTACAAAT 1705
486 ThrSer.....ThrIleAsnAsnLeuValIleAs 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1706 GAGCTAATAATACTGACGCTGCTATCACTTAACAAGCTTGTAATCA 1755
495 nIleSerSerIleAspGlyAlaLysLysAlaLysIleGluThrLysAla 512
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1756 TCTGATTCCTTGGATGAGCTAAAGCGCTGCTTAATGTCAGAGTA 1805
512 hrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeuAspProThr 528
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1806 CCAATGAGCTCTCATATATCCGAACTTTAGAGACTTGTAAGAAACTCT 1855

```

```

539 GlyThrPheTyrGluAsnHisSerLeuArgPro.....GlnSerTyr 543
      ::::::::::::::::::::|||
1856 CAAGATTGGTGTGACACACCGGATGTTAATAAGATTTCACACAGT 1905
543 rasPleuGluLeuLysAla...SerGlyThrValThrSerThrAlaVal 559
      |||||||
1906 TCCGATTTTAGACTCAAGCGACTCAAAATGTAAGTACACCTACGAGCT 1955
559 alThrProAspProIleMetGlyGluLysPheHisTyrGlyTyrGlnGly 575
      ::::::::::::::::::::|||
1956 TCAGTCTGGCCACAACGGCTATACGACATCTCCCTATGGGTATCAAGA 2005
576 ThrTrpGlyProIleValTrpGlyThrGlyAlaSerThrThrAlaThrPh 592
      |||||
2006 ACTTGGAGATTACCATAGACACGACA.....ACCCATACGGTACACAG 2049
592 eAsnTrpThrLysThrGlyTyrIleProAsnProGluArgIleGlySerL 609
      |||||
2050 AAATTGGAAAAAACCGGTTATCTTCATCCATCCGAGCGCTTGCTCCCC 2099
609 euValProAsnSerLeuThrPAsnAlaPheIleAspIleSerSerLeuHis 625
      ||::::|
2100 TCATTCCCTAATAGCCTATGGGCAACGTCATAGATTTCAGAGCTTAGT 2149
626 TyrIleuMetGluThrAlaAsnGlnGlyLeuGlnGlyAspArgAlaPheTr 642
      ::::|
2150 CAAGCGTCAGCAGCTGATGCGGAGATGTCCTGGG...AAGCACTAG 2196
642 pCysAlaGlyLeuSerAsnPhePheHisLysAspSerThrLysThrArg 659
      ::::::|
2197 CATCACAGAAATTACAAATTTCTTCATGCAATCATACCGGTGATGCA 2246
659 rGglyPheArgHisLeuSerGlyGlyTyrValIleGlyGlyAsnLeuHis 675
      ||::::|
2247 GCAGGTACCGCATATGGGTGAGGCTCATCATCAATCACTACACAGC 2296
676 ThrCysSerAspLysIleLeuSerAlaAlaPheCysGlnLeuPheGlyAr 692
      |||
2297 ATCACTCCGATGCTGCGTTAAGTCTAGTTTGGACAGCTGTTACAA 2346
692 gAspArgAspTyrPheValAlaLysAsnGlnGlyThrValTrpGlyTyr 709
      ::::::|
2347 ATCTAAGGATTACTCTCGTAGTCAAGCTCATCTTACGTTATTTCCGTA 2396
709 hrLeuTyrTyrGlnHisAsnGlnThrTyrIleSerLeuProCysLysLeu 725
      ||::::|
2397 CAGTATACTCTAATCATCACCAGCTCTGTGTTGA..... 2431
726 ArgProCysSerLeuSerTyrValProThrGluIleProValLeuPheSe 742
      ::::::|
2432 .....TCATCGAGATCTTCTC 2448
742 rGlyAsnLeuSer.....TyrThrHisThrAspAsnAspLeuLysT 756
      |||||
2449 AGGAGGCACTCTCGAGTTACCTATACCCGATGAGAAAGTAAGA 2498
756 hrLysTyrThrThrTyrProThrValLysGlySerTrpGlyAsnAspSer 772
      ||::::|
2499 CTTCATATACAAATTCCTAAAGGCGCTGCTGTGAGTAACAATTGC 2548
773 PheAlaLeuGluPheGlyArgAlaProIleCysLeuAspGluSerAl 789
      ::|
2549 TGGTTAGGAGACTCGAAGGAACTTCCTCATCTCTCTTCGCAAT 2598
789 aLeu...PheGlnGlnTyrMetProPheMetLysLeuGlnPheValTyr 805
      |||
2599 CTTAACCTCAAGCAGATCATTCCTTTGTAAGAAGTGAAGTTCTTACG 2648
805 laHisGlnGlnGlyPheLysGlnGlnGlyThrGluAlaArgGluPheGly 821
      |||
2649 CGACTCATGGGGCATCCAGAAATAATACCCCGAGGAGGATTTTGA 2698
822 SerSerArgLeuValAsnLeuAlaLeuProIleGlyIleArgPheAspLys 838
      ::::::|
2699 CACGGTCATCTACTCAACGTTGCGGATTCGGTAGGCGTCCGCTTGGTAA 2748
838 sGluSerAspCysGlnAspAlaThrThrTyrAsnLeuThrLeuGlyTyrThrV 855
      ::|
2749 AAATTTCATATATGACACAGATTTTTCACCTATATATGATGCTATGCTC 2798
855 alAspLeuValArgSerAsnProAspCysThrThrThrThrLeuArgIleSer 871
      ||::::|
2799 CTGATGCTATCGTCACAAATCCGATTCGATAGCATACGACATTAATTAAT 2848
872 GlyAspSerTrpLysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuVal 888
      |||
2849 GGACTTACGTGACCTCTATAGGAAATATCTAACCAAGAGTACTTTGCT 2898
888 lLeuArgAlaGlyAsnHisPheCysPheAsnSerAsnPheGluAlaPheS 905
      ::::::|
2899 AGTACAGACATCCAGCCATCTCAGTAATGATGTTCTAGACATCTTGC 2948
905 eArgLysSerPheGlnLeuArgGlySerSerArgAsnTyrAsnValAsp 921
      ::::::|
2949 GGCACGTGTGATGATATTCGCAAGAACCTCCGTAATAATACTTAGAT 2998
922 LeuGlyAlaLysTyrGlnPhe 928
      ::::::|
2999 ATAGGAAGCAAAATTACGATTT 3019

```


Fri Nov 24 13:50:05 2000

us-09-428-122-2.rng

Page 33

OM of: US-09-428-122-2 to: Issued_Patents_NA.* out_format : pfs

Date: Nov 20, 2000 4:37 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame-p2n.model -DEV=x1p
-O/cgn2_1/USPTO.spool/US09428122/runal.14112000.163316.279233/app_query.fasta_1.999
-DB=Issued_Patents_NA -QFM=fastp -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-FCGAPOP=4.500 -FCGAPEXT=0.050 -FCGAPOP=10.000 -FCGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -FCGAPOP=10.000 -FCGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANSNUM=40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09428122.ecgn1_1.67 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-MAIT -THREADS=1
```

Search information block:

```
Query: US-09-428-122-2
Query length: 928
Database: Issued_Patents_NA.*
Database sequences: 262060
Database length: 75620727
Search time (sec): 82.080000
```

score list:

Sequence	Strd Orig	zScore	EScore	len	1	Documentation
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-7 +		271.50	432.62	1.4e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-302-832-7 +		271.50	432.62	1.4e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-530-198-7 +		271.50	432.62	1.4e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-469-880-7 +		271.50	432.62	1.4e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-728-470-7 +		271.50	431.62	1.4e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-617-697-7 +		271.50	431.15	1.7e-16	4287	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-302-832-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-530-198-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-469-880-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-728-470-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-617-697-8 +		270.00	428.74	2.3e-16	4702	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-194-290-6 +		244.50	388.06	4.3e-14	3300	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-614-377A-6 +		244.50	388.06	4.3e-14	3300	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-302-832-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-530-198-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-469-880-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-728-470-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-617-697-5 +		241.50	369.25	4.8e-13	9171	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-302-832-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-530-198-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-469-880-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-728-470-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-617-697-1 +		240.50	375.17	2.2e-13	5116	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-302-832-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-530-198-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-469-880-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-728-470-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-617-697-6 +		223.00	336.14	3.3e-11	9323	
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-409-995-3 +		202.00	305.20	1.8e-09	5738	

```
/cgn2_6/prodata/2/ina/6_COMB.seq:US-08-685-467-3 + 202.00 305.20 1.8e-09 5738
/cgn2_6/prodata/2/ina/6_COMB.seq:US-08-808-599A-23 + 201.50 310.89 8.5e-10 3483
/cgn2_6/prodata/2/ina/5B_COMB.seq:US-08-478-370-1 + 201.00 300.85 3.1e-09 6973
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-317-572A-1 + 200.50 313.35 6.2e-10 2524
/cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-439-818A-1 + 200.50 313.36 6.2e-10 2524
```

```
seq.name: /cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-7
```

seq-documentation_block:

```
; Sequence 7, Application US/08038682
; Patent No. 5549897
;
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
```

COMPUTER READABLE FORM:

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038.682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22.651
REFERENCE/DOCKET NUMBER: 1038-293
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
```

```
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
```

```
US-08-038-682-7
```

alignment_scores:

```
Quality: 271.50 Length: 648
Ratio: 0.828 Gaps: 29
Percent Similarity: 50.617 Percent Identity: 22.531
```

alignment_block:

```
US-09-428-122-2 x US-08-038-682-7 ..
```

```
Align seg 1/1 to: US-08-038-682-7 from: 1 to: 4287
```

```
23 ThrGltValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG1 39
```

```
1955 ACAATATTACCTTTATATGATGACGGGCTATTGACACAAATGGCGCTTC 2004
```

```
39 yAsnPheSerValArgLuser..... 46
```

```
2005 AACCATTCATCCATGCGAGAGGGGCTAAATTAAGATATCAATAACA 2054
```

```
47 .....GlnGluAspAlaGlyThrThyr.....Leu 55
```

```
2055 CCAGTGCTTAATATTATACCAACCACTGTGATGACCACTTACCGCACCACT 2104
```

[illegible]

```

seq.documentation_block:
: Sequence 7, Application US/08302832
: Patent No. 5603938
:
GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoenmaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
:
CLASSIFICATION: 435
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US pct/us93/02166
: FILING DATE: 16-MAR-1993
:
ATTORNEY/AGENT INFORMATION:
: NAME: Beirstlesser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-404
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
:
INFORMATION FOR SEQ ID NO: 7:
:
SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
MOLECULE TYPE: DNA (genomic)
:
US-08-302-832-7
:
alignment_scores:
: Quality: 271.50 Length: 648
: Ratio: 0.828 Gaps: 29
: Percent Similarity: 50.617 Percent Identity: 22.531
:
alignment_block:
US-09-428-122-2 x US-08-302-832-7 ..
:
Align seg 1/1 to: US-08-302-832-7 from: 1 to: 4287
:
23 ThrGluThValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG1 39
:
1955 ACAAATTTACACTTTTATGTAGCGCGCTCATTTGACACAAATGCGCTTC 204
:
39 yasnphSerValArgGluSer..... 46
:
2005 AANACATTTCCATTCGCCAGAGGAGGCGCTAATTTAAAGATATCAATPACA 2054
:
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
:
2055 CCAGTAGCTTAATATATTACACCAACACTGTGATCCACTTACCCGACACAT 2104
:
56 PheTysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAlaI 72
:
2105 ATTAAGGCAATATATCC.....AACCAATCAGGTGATTTGTGATATTAT 2148
:

```

72 ethrlylsSecys.....pheasnant 80
| | | | |
2149 TGTATAAAAAGGACGCGCTGAATTCAAATTGGCGCAATATCTCAAA 2195
| | | | |
80 hrlsGlsAspleuthrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
| | | | |
2199 AAGAGCGAATCTCACAATTTCTTCGTATTAAGTAAATATATTCACATCAG 2246
| | | | |
97 ..ThValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerya 112
| | | | |
2249 ATACACATCAAGACGCGGTGAAGGGGGCTGTGATTAAGTAGAGGC 2296
| | | | |
112 lValAspLySerThr..... 118
| | | | |
2299 AGAAATGCTACCTCACTATTCAAACCAAGCTAAATTAATTGGCAGAG 2348
| | | | |
119PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
| | | | |
2349 ACCTAATATTTTCAGGCTTTAATAAGCAGAAATACAGCTTAAATATGC 2398
| | | | |
133 SerSerIleThrThrGlyLyGlyAlaValAsnSerSeryaThrGlySerye 149
| | | | |
2399 AGGATTTTACCTATTGGC.....ATGCTACGGGTGTAAATGC 2436
| | | | |
149 uSerLeuThrLys..... 153
| | | | |
2437 TGAATCTAAAAAGATGACATTGTGACAAAGTTAAAGATTCAAAATCTCGA 2486
| | | | |
154AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
| | | | |
2487 CTGACGGTCACATGTAACACATA.....AAATGGGAAGTGAAGAAACGCTC 2530
| | | | |
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThme 183
| | | | |
2531 AATGTAAGTAC...AATGCTGTATGATTAACACACCGGTTTAACAT 2577
| | | | |
183 tSerAla.....LeuPheSerGluAsnThrSerSeryLys..... 195
| | | | |
2578 TTCGGCAAGATGTAGACGTTAAACAAATACGTTACCTCCCAAGAACAA 2627
| | | | |
196GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
| | | | |
2628 TAAATATCTCTCCCGACGACGAGAAATGTAAACACCAAGAAAGGCACACT 2677
| | | | |
207 lIeThrGlyAsnGlyGlyValuValSerPheSerAspAsnThrSerSerys 223
| | | | |
2678 AATCATGTCAACACACAGCAGCGTGAAGTAACTGCTCAAAATGTGACAT 2727
| | | | |
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnSna 240
| | | | |
2728 TAAAGGCAC.....ATTACCTGCAAAATGTAAACACTGACACACATCAG 2771
| | | | |
240 lAluValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 256
| | | | |
2772 AAAATCTGTTTACCACAGAGATGCTGTCAATTATGCACACAGCGGACCA 2821
| | | | |
257ThGlyAspSerGlyGlyAlaIleCysAl 267
| | | | |
2822 GTAACATTAGTACAAAAACGCGGATTTTAAAGGTGA.....AT 2862
| | | | |
267 alyrlyThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
| | | | |
2863 TCAATCACTTCGGTATATGTAAATATTACAGCGAGCGGCAT...ACAC 2909
| | | | |
284 euleuPheSerAsnThr..... 290
| | | | |
2910 TTAAGGTAAATATCTACTGCTCAAGATGTAAACAGTAAACAGCGGATGCA 2959
| | | | |
291SerThrThrAlaGlyGlyAlaIleLysValLys..... 301
| | | | |
3960 GGAGCCTTGACACACTACGACGCGCTCAACCACTTATGTCGACACACAGCA 3009

```

302 .....LysLeuGluLeu 306
3010 TCGAATATTACACCAAAAGCATATCAACGGTAAGTTGATCA 3059
306 1aSerGlyLeuThrLeuPheSerArgAsnSerValasnGlyThr 322
3060 GCTCCGGCTGTAACTGTTGGCAACGGAGCAACTTGTCTAGT 3109
323 AlaProGlyGlyAlaIleAlaIle...GluAspSerGlyGluLeu.. 337
3110 AATATTTCAGTAACACTGTTACTTACTGCGGATAGCGTAATTAATAC 3159
338 .....SerLeuSer 341
3160 CTCACAGTAGTCTTACATTAATGGAGCTAATAGTACACCACTCA 3209
341 1aAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
3210 GCCATTCAGCGCATATTGAGGTACATTTCTGGTAATACGTAATGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
3260 ACA.....GCAAGCACTGTGATTACTTACTTGAATAATAGTCAA 3300
372 SerThrAlaLeuArgSerAlaIleAlaGlyArgAlaIleTyrPheTyrAsp 389
3301 AGTGAACCGCAAAATGGAGCTGCA..... 3325
389 rollerThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIlePhe 422
3325 ..... 3325
422 rGlyGluLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThr 439
3326 .....ACCTTAAGTCTGTAATCAGCAAAATTAAACCA 3356
439 eRlyLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
3357 CCCAAACAGCGCTTACATTACTCAAGCAATGTCAGCAACTCTTACA 3406
454 LeuLysHis.....GlyValThrIle 460
3407 GCCAAGATAGCAGTATCGAGAAACATTATGCTGTAATGTGACGTT 3456
460 uGlnThrGln...AlaPheThrGlnGlnAlaAspSerArgLeuGluMet 476
3457 AATATCCAGAGCGACTTAACTACTACAGGAGGATTCAAAGATTACGCA 3506
476 sPValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
3507 CCAGTGTACCTTAACATCAATGCAAAAGATGCCAAATTAGATGTGCT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
3557 GCATCGAGTGCAGCAGTAACTAGTAATGCAACTACGACAGGCTCTGCG 3606
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSer 520
3607 TAACTG.....ACGGCAAAACCTCAAGCAGCGTGAATATCAACG 3647
520 1yThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsn 534
3648 GGGATTTAACACATAAATGGTTAAATATCATTTTCGAAAT 3691

```

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:us-08-530-198-7

seq_documentation_block:
; Sequence 7, Application US/08530198
; Patent No. 5869065

```

? GENERAL INFORMATION:
? APPLICANT: BARENKAMP, STEPHEN J
? APPLICANT: ST. GEME ITI, JOSEPH W
? TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
? TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESS: Shoemaker and Matiere, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/530,198
? FILING DATE: 13-DEC-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: BERKSTRESSER, JERRY W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: JWB-1186
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4287 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-530-198-7

```

alignment_scores:
Quality: 271.50 Length: 648
Ratio: 0.828 Gaps: 29
Percent Similarity: 50.617 Percent Identity: 22.531

alignment_block:
US-09-428-122-2 x US-08-530-198-7 ..
Align seg 1/1 to: US-08-530-198-7 from: 1 to: 4287

```

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG 39
1955 ACAATTAACACTTTAATGTAAGCGGCTCATTTGACACAAATGCGCTTC 2004
39 yAsnPheSerValArgGluSer..... 46
2005 AAACATTTCATGCGAGAGAGGAGGCTAAATTTAAAGATATCATATACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
2055 CCAGTACCTTAATATACCAACCACTGATACCATTCACGACCAAT 2104
56 PheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAla 72
2105 ATAAAAAGCAATATATCC.....AACAAATCAGGTGATTGAAATATAT 2148
72 eThrLysSerGys.....PheAsnAsnT 80
2149 TGATAAAAAAGCGAGCGCTGAATCCAAATTTGGCGCAATATCTCACAAA 2198
80 hLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
2199 AAGAAGCAATCTCACAATTTCTTGTGATAAGTAATATTTACCAATCAG 2248

```

```

97  ...ThrValaspAlaGlyThrValAlaGlyAlaAlaValAsnSerSera 112
112  ValAspLysSerThrThr.....SerLeuSera 118
2299 AGAAATGCTACCTACTATTCAACCAAGATTAAATGGCAGAG 2348
119  ....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
2349 ACCTAAATATTTCAGCCTTATATTAAGCAGAAATTACAGCTAAATATGCC 2398
133 SerSerIleThrThrGlyGlyGlyAlaValSerSerserThrGlySerLe 149
2399 AGTATTTAACTATTGGC.....AATGCTAGCGGTGGTAAATGC 2436
149 SerLeuThrLys..... 153
2437 TGATGCTAAATAAGTACTTTGACAAAGTTAAAGATTCAAAATCTCGA 2486
154  ....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
2487 CTGACGGTCAACAATGTAACTA....AATAGCGAAGTGAATAACCTCT 2530
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrme 183
2531 AATGTGTAGTAC...AATGCTGTATATGATACAGCACCGGTTTAAACCAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
2578 TTCGCAAAAGATGTACGGTAAACAATTAAGTAACTCCACAGACAA 2627
196  ....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
2628 TAAATATCTCTGCCGACAGCAAGAAATGTAAACCAAGAGCAGCACT 2677
207 IleThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSera 223
2678 ATCATATGCAACACAGCAGCGTAGAGTACTGCTCAAAATGGTACAAAT 2727
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsn 240
2728 TAAAGGCAC....ATTAAGTGTAAATGTAAATGTACAGGACGACAG 2771
240 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 256
2772 AAATCTTTGTACACAGAGAAATGCTGTATATGCAACCGCAGCAGCA 2821
257  ....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
2822 GTAACATGTAGTACAAACAGAGGATATTAAGGTGA.....AT 2862
267 aTyrlYsThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
2863 TGATCAACTTCGCGTAATGTAAATATTAACAGGACGCGCAAT...ACAC 2909
284 euleuPheSerAsnThr..... 290
2910 TTAAGTAAGTATATCACTGTCAGATGTAAACAGTAAACGCGATGCA 2959
291  ....SerThrThrAlaGlyGlyAlaIleTyValLys..... 301
2960 GGAGCCTTGACACTACAGCAGGCTCAACCATTAAGTCCGACACAGGCA 3009
302  ....LysLeuGluLeuA 306
3010 TCGAAATATTACAAACAAACAGGTGATATCAACGTAAGTTGAATCCA 3059
306 lAserGlyGlyLeuThrLeuPheSerThrGlnSerValAsnGlyGlyThr 322
3060 GCTCCGGCTGTGTACACTGTGTGCAACTGGACCACTCTTGTCTAGGT 3109

```

```

323 AlaProLysGlyGlyAlaIleAlaIle...GluAspSerGlyGluLeu... 337
3110 AATATTTCAGTACACACTGTACTATTACTGCGGATAGCGGTAAATTAAC 3159
338  ....SerLeuSera 341
3160 CTCACACATAGCTTACATTAATGCGACTAATAGTACACCTCA 3209
341 lAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
3210 GCCAATCAGCGCATATGTGAAGTACAAATTCGTGTAATACAGTAATGT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
3260 ACA.....GCAAGCACGTGGTATTAATGAAATAGTGCATA 3300
372 smetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyPheTyAsp 389
3301 AGTTGAAGCGAAATAAGAGCTCA..... 3325
389 rollerThrThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
3325  .... 3325
406 GluThrProAlaAspSerAlaLeuGlnTyThrGlyAsnIleIlePheThr 422
3325  .... 3325
422 rGlyGlyLysLeuSerGlyThrGluAlaAlaAspSerLysAsnLeuThrS 439
3326  ....ACCTTAAGTCTGTGATATCAGCAATTAATACCA 3356
439 erLysLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
3357 CCCAAACAGCTCTTACATTACTTACCAAGCAATGTCCAGCAACTCTTACA 3406
454 LeuLysHis.....GlyValThrLe 460
3407 GCCAAGATAGACATGTCCAGCAAAACATTATGCTGCTAATGTACGCTT 3456
460 uGlnThrGln...AlaPheThrGlnGlnAlaAspSerArgLeuGlnMetA 476
3457 AATACACACAGCAGCTTACTACTACAGGAGGATTAAGATTAAACGCA 3506
476 spValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
3507 CCAGTGTACTCTTAACAAATCAATGCAAAAGATGCCAAATTAGATGCTGT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 503
3557 GCATCAGGTGACCGCACAGTACTAATGCACTAACGCAAGTGGCTCTGG 3606
503 sLysAlaLysIleGlyThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
3607 TAACGG.....ACTGCGAAACCTCAAGACGCGGATATACACCG 3647
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyGluAsn 534
3648 GCGATTTAACACATTAATGGCTTAATATATCATTTCCGAAAT 3691

```

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-469-880-7

seq_documentation_block:

Sequence 7, Application US/08469880

Patent No. 5876733

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J.

TITLE OF INVENTION: High Molecular weight Surface Proteins

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Maltare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,880
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: US-08-469-880-7

```

```

alignment_scores:
  Quality: 271.50      Length: 648
  Ratio: 0.828        Caps: 29
  Percent Similarity: 50.617  Percent Identity: 22.531

```

alignment_block:

US-09-428-122-2 x US-08-469-880-7 ..

Align seg 1/1 to: US-08-469-880-7 from: 1 to: 4287

```

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG1 39
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1955 ACAAAATTACACTTTTAATGTACGCCGCTCATTTGACAAACAATGCGCTTC 2004
39 yAsnPheSerValArgGluSer..... 46
: : : : : : : : : : : : : : : : : : : : : : : :
2005 AAACATTTTCATTGCGCAAGAGGGGCTAAATTAAAGATATCAATAACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
2055 CCAGTAGCTTAATATATACCAACAACAGCTGATACCACTTACCGCACTT 2104
56 PheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAlaI1 72
: : : : : : : : : : : : : : : : : : : : : : : :
2105 ATAAAGCAATATATATCC.....AACAAATCAGGTGATTGAATATATAT 2148
72 eThrLysSerCys.....PheAsnAsnT 80
||||| : : : : : : : : : : : : : : : : : : : : : : :
2149 TGATATAAAAAAGCAGCGCTGAATCCAAATTGGCGGCAATATCCACAAA 2198
80 hrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
: : : : : : : : : : : : : : : : : : : : : : : :
2199 AAGAAGCAATCTCACAATTTCTCTGATTAAGTAATATATACCAATCAG 2248

```

```

97 ...ThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
||||| : : : : : : : : : : : : : : : : : : : : : : :
2249 ATACAAATCAAAACAGCGCTTGAGGGGGCGTCTGATTCAGATGAGAGC 2298
112 ValAspLysSerThrThr..... 118
: : : : : : : : : : : : : : : : : : : : : : :
2299 AGAAATGCTAACTACACTATTCACCAACCAAGAGTTAAATTCGACAGAG 2348
119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
: : : : : : : : : : : : : : : : : : : : : : :
2349 ACCTAAATATTTCAGGCTTTAATAAAGCAAGAAATACAGCTAAATAATGC 2398
133 SerSerIleThrThrGlyLysGlyValAlaLysCysSerThrGlySerLe 149
||||| : : : : : : : : : : : : : : : : : : : : : : :
2399 AGTGAATTAACATATGCG.....AATGCTAGCGGTGGTGAATGC 2436
149 uSerLeuThrLys..... 153
: : : : : : : : : : : : : : : : : : : : : : :
2437 TGATGCTAAAAAAGTGAAGTTTGACAAAGTTAAAGATTCAAAATCTCGA 2486
154 .....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
||||| : : : : : : : : : : : : : : : : : : : : : : :
2487 CTGAGCGTCACAAATGTAACTA.....AATAGCGAAGTGAATAACCTCT 2530
167 AsnGlyAlaAlaIleThrAlaLysThrLeuSerLeuThrGlyThrPhe 183
||||| : : : : : : : : : : : : : : : : : : : : : : :
2531 AATGTTACTAGC...AATGCTGTAATGATACAGCCGCTTAACCAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
||||| : : : : : : : : : : : : : : : : : : : : : : :
2578 TTCCGCAAAAGATTAACGGTAAACAAATAAGCTTACCTCCCAAGACAA 2627
196 .....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
||||| : : : : : : : : : : : : : : : : : : : : : : :
2628 TAAATATCTCTGCCGACAGCAAGTAATGTAACACCAAGAGCACACT 2677
207 IleThrGlyAsnGlnGlyLysValSerPheSerAspAsnThrSerSer 223
||||| : : : : : : : : : : : : : : : : : : : : : : :
2678 ATCAATGCAACACAGCGCACGCGTGAAGTACGCTCAAAATGTATCAAT 2727
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsn 240
||||| : : : : : : : : : : : : : : : : : : : : : : :
2728 TAAAGCGAAC.....ATTACCTCGCAAAATGTACACAGTACGACACAG 2771
240 LeuLysValSerPheIleAspAsnLysValThrGlyLysSerSerThr 256
||||| : : : : : : : : : : : : : : : : : : : : : : :
2772 AAATCTTTTACACAGAGAAATGCTGATTAATGACCAACGCGGCACA 2821
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
||||| : : : : : : : : : : : : : : : : : : : : : : :
2822 GTAACATTAGTACAAAAACAGGAGATATTAAGTGA.....AT 2862
267 eTryLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
||||| : : : : : : : : : : : : : : : : : : : : : : :
2863 TGAATCACTTCGGTAAATGTAATATATACAGCAGCGCAAT...ACAC 2909
284 euleuPheSerAsnAsnThr..... 290
||||| : : : : : : : : : : : : : : : : : : : : : : :
2910 TTAAGTAACTAATATCACTGCTCAAGTGTACAGTACAGCGGATGCA 2959
291 .....SerThrAlaGlyAlaIleThrValLys..... 301
||||| : : : : : : : : : : : : : : : : : : : : : : :
2960 GGAGCGCTGACAACTACAGCAGCGCTCAACCTTGTGCGAACACAGCGCA 3009
302 .....LysLeuGluLeuA 306
||||| : : : : : : : : : : : : : : : : : : : : : : :
3010 TGCAAAATTTCAACCAACAGGTGATATCAACGGTAAAGTTGATTCGA 3059
306 laSerGlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 322
||||| : : : : : : : : : : : : : : : : : : : : : : :
3060 GCTCGCGCTGTGAACACTTGTGCAACACTGGACACACTCTTGTGTAGGT 3109

```



```

333 AlaProlysglyAlaIleAlaIle...GluAspSerGlyGluLeu.. 337
      ::::::::::::::::::::|||
3310 AATATTGAGAACACTGTACTATCTACTGGGATACGGTAATTAAC 3159
      ::::::::::::::::::::|||
338 .....SerLeuSera 341
3160 CTCACAGTACAGTCTACAAATTAATGAGACATAATAGTACACACCTCAA 3209
341 laasPserGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
      ::::::::::::::::::::|||
3210 GCCAATCAGCGGATATGAGGTACAAATTCGTGTAATACAGTAATATGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAlaLy 372
      ::::::::::::::::::::|||
3260 ACA.....GCCAACGCTGGTGATTTACTATTGAGAAATAGTGCAA 3300
372 smethrAlaLeuArgSerAlaIleGlyArgAlaIleThrPheThrAsp 389
      ::::::::::::::::::::|||
3301 AGTTGACGCAAAATGAGCTGCA..... 3325
389 rolleThrThrGlySerSerThrThrValThrAspValLeuValAsn 405
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnThrThrGlyAsnIleIlePheTh 422
3325 ..... 3325
422 rGlyGluLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThrs 439
      ::::::::::::::::::::|||
3326 .....ACCTTACCTGCTGATCAGCAACGCAATTAACCA 3356
439 erLysLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
      ::::::::::::::::::::|||
3357 CCCAAACAGCGCTAGCATTAACCTCAACCAATGTCAGACAACTCTTACA 3406
454 LeuLysHis.....GlyValThrIle 460
3407 GCCAAGGATACAGATTCGACGAAACATTAATGCTGTAATGTGACGTT 3456
460 uGlnThrGln..AlaPheThrGlnAlaAlaAspSerArgLeuGluMeta 476
      ::::::::::::::::::::|||
3457 AAATACCCACAGCACTTAACCTTACACTACAGGAGTCAAGATTAACGCA 3506
476 spValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
      ::::::::::::::::::::|||
3507 CCAGTGTGATCCTTAACATCAATGCAAAAGATGCCAAATAGATGCTGCT 3556
487 serThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLy 503
      ::::::::::::::::::::|||
3557 GCATCAGGTGACCGCACAGTAGTAATCAACTACGCAAGTGCCTCG 3606
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
      ::::::::::::::::::::|||
3607 TAAGGTG.....ACTGCGAAACCTCAAGCAGCTGAAATACACCG 3647
520 LysThrIleThrLeuLeuAspProThrGlyThrPheThrGlyAsn 534
3648 GGGATTTAAACACATTAATAGGTTAATATCATTTGCGAAAT 3691
seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:US-08-728-470-7
seq_documentation block:
: Sequence 7, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SPOUNCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ. ID NO.: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-7

alignment_scores:
  Quality: 271.50      Length: 648
  Ratio: 0.828        Gaps: 29
  Percent Similarity: 50.617  Percent Identity: 22.531

alignment_block:
  US-09-428-122-2 x US-08-728-470-7 ..

Align seg 1/1 to: US-08-728-470-7 from: 1 to: 4287

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG 39
      ::::::::::::::::::::|||
1955 ACAAAATTACCTTTAATGTAGCGGCTCATTTGACAAACATGCGCTTC 2004
39 YAsnPheSerValArgLysSer..... 46
      ::::::::::::::::::::|||
2005 AAACATTTCCATTGCTCCAGAGAGGGGCTAAATTAAGATATCAATACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
      ::::::::::::::::::::|||
2055 CCACTACCTTAATATATACCAACCACTGTGATACCACTTACCCACCAT 2104
56 PheLysGlyAsnValThrIleGluAsnIleProGlyThrThrAlaIle 72
      ::::::::::::::::::::|||
2105 ATAAAGCAATATATTC.....AACAAATCAGGTGATTTGAATATTAT 2148
72 eThrLysSerCys.....PheAsnAsnT 80
      ::::::::::::::::::::|||
2149 TGAATAAAAGGAGCGCTGAATTCGAATTTGGCGGCAATATCTCACAA 2198
80 hLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
      ::::::::::::::::::::|||
2199 AAGAAGCAATCTCACAAATTTCTTGATTAAGTAATATTAACCAATCAG 2248

```

```

97  ...ThrValAspIaGlyThrValaIaGlyAlaValaIaSerSerVa 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2249 ARAACAAATCAAAAGCAGCGCTTAAGGGGGCGCTTCTGATTCAGTGAAGGC 2298
    112 lValAspIysSerThrThr..... 118
    :   :   :   :   :   :   :   :   :   :   :   :   :
2299 AGAAAATGCTAACTAACTTCAACCAAGACTTAAATTTGGCAGAG 2348
    119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
    :   :   :   :   :   :   :   :   :   :   :   :   :
2349 ACCTAAATATTCAGCGCTTATTAAGCAGAAATTACAGCTAAATAAGGC 2398
    133 SerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerIle 149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2399 AGTGTAACTATTGGC.....AATCTAGCCGCTGGTAAATGC 2436
    149 uSerLeuThrLys..... 153
    :   :   :   :   :   :   :   :   :   :   :   :   :
2437 TGATGCTAAAAAGTGAAGCTTTGACAAAGCTTAAAGATTCAAAATCTCGA 2486
    154 .....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2487 CTGACGGCTCAAAATGTAACTA.....AATGCGAAGTGAAGAAAGCTCT 2530
    167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrme 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2531 AATGCTAGTGAAC...AATGCTGTAATGATTAACAGCAGCGCTTAAACAT 2577
    183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2578 TTCGCGAAAAGATGTAAAGGTAAACAATAAGTAACTCTCCCAAGACAA 2627
    196 .....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2628 TAAATCTCTGCGGAGCAGGAATGTAAACAACCAAGAGGCGCAACT 2677
    207 lIleThrGlyAsnGlnGlyuValSerPheSerAspAsnThrSerSerAs 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2678 ATCAATGCAACCAAGCAGCGCTGAAGTAAGTCTCAAAATGGTACAT 2727
    223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsn 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2728 TAAAGGCAC...ATTACCTCGCAAAATGTAACTAGTACAGACACAG 2771
    240 lAlaValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2772 AAAATCTGTGTACACAGAAATGTGTCAATTATGACACAGCGGCACA 2821
    257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2822 GTAAACATTACTACAAAACAGGGGATATTAAGTGA.....AT 2862
    267 atyLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2863 TGAATCAACTCCGCTAATGTAATATATTAACAGCAGCGGCAT...ACAC 2909
    284 euleuPheSerAsnAsnThr..... 290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2910 TTAAGTAAGTAATATCACTGCTCAAGATGTAAACAGTAAACAGCGATGCA 2959
    291 .....SerThrAlaGlyGlyAlaIleThrValLys..... 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2960 GGAGCTTGACACACTACAGAGGCTCAACCATTAAGTGCAGCAACAGCAA 3009
    302 .....LysLeuGlnLeuA 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3010 TGCAAAATATTAACCAAAACAGTGATATCAACGGTAAAGTTGAATGCA 3059
    306 laseGlyGlyLeuThrIlePheSerThrAsnSerValAsnGlyLysThr 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3060 GCTCCGGCTCTGTACACTTGTTCGAACGTGAGCAACTCTTCTGTAGT 3109

```

```

323 AlaProlyGlyGlyAlaIleAlaIle...GluAspSerGlyLeu... 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3110 AATATTTCAGTAAACACACTTACTTACTTACTGCGATAGCGGTAAATTAC 3159
    338 .....SerLeuSerA 341
    :   :   :   :   :   :   :   :   :   :   :   :   :
3160 CTCCACACTAGTCTCTCAATTAATGGACTAATAGTAAACCACTCA 3209
    341 lAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3210 GCCAATCAGCGGATTAAGAGTAAAGTACAACTTCTGCTAATACAGTAATGTT 3259
    356 ThrThrProGlyThrAsnArgSerIleAspLeuGlyThrSerAla 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3260 ACA.....GCAACACTGCTGATTTACTTATGGAATAATAGTCANA 3300
    372 sMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrrPheTyrrAsp 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3301 AGTTGAAGCGAAAATGGAAGCTCA..... 3325
    389 rollerThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
    3325 ..... 3325
    406 GluThrProAlaAspSerAlaLeuGlnTyrrGlyAsnIleIlePheTh 422
    3325 ..... 3325
    422 rGlyGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThr 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3326 .....ACCTTAACCTGCTGAATCAAGCAATTAACCA 3356
    439 eryLysLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3357 CCCAAACAGCGCTGTGACTTACCTCAAGCAATGTCAGACAACTTACA 3406
    454 LeuLysHis.....GlyValIleThrLe 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3407 GCCAAGGATAGCAGTATGCGAAGAACTTATGCTCTAATGTGAGCTT 3456
    460 uGlnIleThrGln...AlaPheThrGlnGlnAlaAspSerArgLeuGlnMet 476
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3457 AATATCCACAGCGCACTTAACTACTACAGGAGTAAAGATTAAACCA 3506
    476 sValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3507 CCAGTGTACCTTAAACATCAATCAAAAGATGCCAAATTAGTGTGCT 3556
    487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3557 GCATCAGGTGACCCACAGTAGTAATGCAACTAAGCAAGTGTGCTGCG 3606
    503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3607 TAAGTGT.....ACTCGAAAAACCTCAAGCAGCGTGAATATCACCG 3647
    520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyrrGluAsn 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3648 GGGATTTAACACATAAATAGGTTAATATATCATTCGGAATAT 3691

```

seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:US-08-617-697-7

seq_documentation_block:

Sequence 7, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4794 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear

```

alignment_scores:

```

Quality: 271.50      Length: 648
Ratio: 0.828         Gaps: 29
Percent Similarity: 50.617   Percent Identity: 22.531

```

alignment_block:

```
US-09-428-122-2 x US-08-617-697-7
```

```
Align seg 1/1 to: US-08-617-697-7 from: 1 to: 4794
```

```

23 ThrGluThrValLeuAspSerSerLasePheAspGlyAsnLysAsnG1 39
||||| : : : : : : : : : : : : : : : : : : : : : : : :
2734 ACAATTCACCTTTTATGTAAGCCGCGTCATTTGACACAATGCGCGTTC 2783
39 yAsnPheSerValArgLysSer..... 46
||||| : : : : : : : : : : : : : : : : : : : : : : :
2784 AAACATTTCCATTGCCAGAGGAGGGGCTAAATTAAAGATATCAATACA 2833
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
||||| : : : : : : : : : : : : : : : : : : : : : : :
2834 CCAGTAGCTTAATATTACCAACACACTGTGATCCACTTACCGCACCAAT 2883
56 PheLysGlyAsnValThrLeuGluAsnLysProGlyThrGlyThrAla11 72
||||| : : : : : : : : : : : : : : : : : : : : : : :
2884 ATAAAGGCAATATATTC.....ANCAATCAGGTGATTTGAATATAT 2927
72 eThrLysSerCys.....PheAsnAsnT 80
||||| : : : : : : : : : : : : : : : : : : : : : : :
2928 TGATTAATAAAAGCGACGCTGAATCCAATTGGCGGCATATCTCACAAA 2977
80 hLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuPheGln 96
||||| : : : : : : : : : : : : : : : : : : : : : : :
2978 AAGAAGGCAATCTCACAATTTCTTGTGATAAGTAATATTCACATCAG 3027
97 ..ThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
||||| : : : : : : : : : : : : : : : : : : : : : : :
3028 ATAACAAATCAAGCAGCGCTTGAAAGGGGGCGTCTGATTCAGTAGAGC 3077

```

```

112 lValAspLysSerThrThr..... 118
: : : : :
3078 AGAAATGCTAACCTTACTTATCAACCAAGACTTAAATTTGCAGAG 3127
119 .....PheLysGlyPheSerSerLeuSerPheLysLeuSerProGly 132
: : : : :
3128 ACCTAATATTTCAGGCTTTATTAACACGAAATTAACGCTAAATATGC 3177
133 SerSerLysThrThrGlyGlyAlaValSerCysSerThrGlySerLe 149
||||| : : : : :
3178 AGTCAATTACTATATGCG.....ATCTAGCGGTGATATGC 3215
149 uSerLeuThrLys..... 153
: : : : :
3216 TGATGCTAAAGTAGACTTTTGACAAGGTTAAGATTCAAAATCTCGA 3265
154 .....AsnValSerLeuPheSerLysAsnPheSerThrAsp 166
||||| : : : : :
3266 CTGACGGTCACAAATGTAACACTA.....ATACGGAAGTGAAAACGCT 3309
167 AsnGlyAlaAlaThrAlaLysThrLeuSerLeuThrGlyThrPhe 183
||||| : : : : :
3310 AATGTAAGTAC.....ATGCTGTAATGATATACAGACCGGTATACCAT 3356
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
||||| : : : : :
3357 TTCGCAAAAGATGTAAACGTTAAACAAATTAACCTCCACCAAGACAA 3406
196 .....GlyGlyAlaAlaGlnThrPheSerAspAlaLeuThr 206
||||| : : : : :
3407 TAAATATCTGCGCCGACAGAGAAATGTACACCAACCAAGGACCAACT 3456
207 lLeThrGlyAsnGlnGlyGlyValSerPheSerAspAsnThrSerSer 223
||||| : : : : :
3457 ATCAATGCAACCAACGAGCGGTGGAAGTACTCTCAAAATGATACAT 3506
223 pSerGlyAlaAlaLeuPheThrGluAlaSerValThrLysSerAsn 240
||||| : : : : :
3507 TAAAGGCAC.....ATTACCTGCAAAAGTAAACAGTGACACACAG 3550
240 lAlysValSerPheLysAsnLysValThrGlyAlaSerSerThr 256
||||| : : : : :
3551 AAAATCTTGTACCAACAGAGATCTGTATTAATGACACAGCGGACCA 3600
257 .....ThrGlyAspMetSerGlyAlaAlaLysAl 267
||||| : : : : :
3601 GTAAACATTAGTACAAAACAGGCGATATTAAGTGGA.....AT 3641
267 aTyLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
||||| : : : : :
3642 TGAATCAACTCCCGGTAAATTAATATTACGAGCGGCGCAAT...ACAC 3668
284 eLeuPheSerAsnAsnThr..... 290
||||| : : : : :
3669 TTAAGTAGTAATATATCACTGCTCAAGATGTAAACAGTACGCGATGCA 3738
291 .....SerThrThrAlaGlyAlaAlaLysValLys..... 301
||||| : : : : :
3739 GGACCTTGACAACTACAGCAGCGTCAACCTATGTCGACACAGCAA 3788
302 .....LysLeuGluLeu 306
||||| : : : : :
3789 TGCAAAATATTCAACAAACAGCGTGAATATCAACGTTAAAGTTGAATCA 3838
306 lAserGlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 322
||||| : : : : :
3839 GCTCGGCTGTGTAACACTGTGTAACAGTGAACACTGCTGTAAGGT 3888
323 AlaProLysGlyAlaAlaLysLeu...GluAspSerGlyGluLeu. 337
||||| : : : : :
3889 AATATTTCAGTAACTGTTACTATTAATGCGGATGCGGTAAATTAAC 3938

```

```

338 .....SerLeuSerA 341
3399 CTCACAGTAGTCTACATTAATGAGACTAATAGTGTACACCACTCAA 3988
341 laasSercIyaspIle.....ValpheLeuGlyasnThrValThrser 355
3989 GCCAATCAGGCGATATGAGGTACAAATTCGTGAATACAGTAATCTT 4038
356 ThrThrProGlyThrAsnArgSerSerIleaspLeuGlyThrSerAlaLy 372
4039 ACA.....GCACGACTGGTGATTTAACTATGTGAAATAGTCGAAA 4079
372 smethrAlaLeuArgSerAlaAlaGlyArgAlaIleTyPheTyrasp 389
4080 AGTTGAGCGAAATAATGAGCTGCA..... 4104
389 roIleThrGlySerSerThrThrValThrAspValLeuLysValasn 405
4104 ..... 4104
406 GluThrProAlaAspSerAlaLeuGlnTyThrGlyAsnIleIepheth 422
4104 ..... 4104
422 rGlyGlyLeuSerIuThrGlyAlaAlaAspSerLysAsnLeuThrs 439
4105 .....ACCTTAACGTCTGAATCAGCAATTAACCA 4135
439 erLysLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
4136 CCCAACAGGCTGTAGCATTACTCAAGCATGTGTAGCAACACTCTTACA 4185
454 LeuLysHis.....GlyValThrIle 460
4186 GCCAAGGATAGCATGTCCAGAAACATTAACTGCTAATGTGACGTT 4235
460 uGlnThrGln.....AlaPheThrGlnGlnAlaAspSerArgLeuGlnMetA 476
4236 AATATACACAGCAGCATTAACTACTACAGGAGATTCAAAGATTACGCAA 4285
476 spValGlyThr...ThrLeuGlnProAlaAsp.....Thr 486
4286 CCAAGTGGTCTTAACAAATCAATGCAAAAGTCCAATTAAGTGTGCT 4335
487 SerThrIleAsnLeuValIleAsnIleSerSerIleaspGlyAlaLy 503
4336 GCATCAGGTGACCGCAGTAGTAATCAACTAACGCAAGTGGCTCTGG 4385
503 slySalAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerg 520
4386 TAACTGTG.....ACTCGAAACCTCAAGCAGCGTGAATATCAACG 4426
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyrgLysAsn 544
4427 GGGATTTAACACAAATAATGCTTAATATCATTTTCGAAAAAT 4470
seq_name: /cgn2_6/prodata/2/lna/5A.COMB.seq:us-08-038-682-8
seq_documentation_block:
; Sequence 8, Application us/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia

```

```

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERSKRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4702 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-038-682-8

```

```

alignment_scores:
Quality: 270.00 Length: 644
Percent Similarity: 49.845 Percent Identity: 22.671

```

```

alignment_block:
US-09-428-122-2 x US-08-038-682-8

```

```

Align seg 1/1 to: US-08-038-682-8 from: 1 to: 4702

```

```

27 LeuAspSerSerAlaSerPheaspGlyAsnLysasnGlyAsnPheserVa 43
2537 CTAAACATCACCGGACCTTTACCAACAAACGTCACCGCAACATTATAT 2586
43 lArgLysSerGlnGlnAspAlaGlyThrThrTyLeuPheLysGly.... 58
2587 AAAACAAGAGTGGTAAACCTCCAGGCGATATATCAATTAAGGTGTT 2636
59 ..AsnValThrLeuGlnAsnIleProGlyThrGlyThrAlaIleThrLys 74
2637 TAAATATCACT...ACTTACGCTCAGGCACTCAAAAACCATTTATTAAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr..... 87
2684 GGAATATATACTTAACGAATAAAGCGACTTAACATCAAGATATTAAACG 2733
88 .....GlyAsn.....GlyAsn 92
2734 CGAGCGCAAAATCCAAATTCGCGCAATATCTCACAATAAAGAGCAATC 2783
92 erLeuLeuPheGln.....ThrValasp 99
2784 TCACAAATTTCTTGATAAAGTAATATTACCAATCAGATACAAATCAA 2833
100 AlaGlyThrValAlaGlyAlaValaValaAsnSerSerValValaLysLys 116
2834 GCAGCGCTTAAGGGCGGCTTCTGATTTCAAGTGAGGCAAAATGCTTAA 2883
116 rThrThr.....PheI 120
2884 CTTAACTATTCAACCAAGATTAATTTGCGAGAGACCTAAATATTT 2933
120 leGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
2934 CAGGCTTATAATAAGCAAAATTACAGCTAAATAATGGCAGTATTAACT 2983

```

```

137 ThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrLy 153
1288 ATGGC.....AATGCTACGGGTGTAATGCTGATGCTAATAA 3021
153 S.....A 154
3022 AGTGAAGCTTTGACAGGTTAAAGATTCAAAAATTCGACTGACGGTCACA 3071
154 snValSerLeuPheSerLysasnPheSerThrAspAsnGlyAla 170
3072 ATGTAACACTA.....AATAGCCAAGTGAAGAAAGCTAATAGTAGAGC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSerAla.... 185
3116 ..AATGCTGTAATGATTAACAGCACCGGTTTAACCATTCGCCAAGA 3162
186 ..LeuPheSerGluAsnThrSerSerLys..... 195
3163 TGTAACGGTAACAAATACGTTACCTCCACAGACAAATTAATCTCTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
3213 CCGCAGCAGGAATGTAACAACCAAGAGCAGCACTATCAATGCAACC 3262
211 GlnGlyLysValSerPheSerAspAsnThrSerSerLysGlyAlaAl 227
3263 ACAGGCGCGGTGAAGTAACTGCTCAAAATGCTACAATTAAAGCAAC. 3310
227 AliePheThrGluAlaSerValThrIleSerAsnAlaLysValSerP 244
3311 ..ATTACCTCGCAAAATGTAACAGTACAGCAGCAAGAAATCTGTGA 3356
244 heIleAspAsnLysValThrGlyAlaSerSerThr..... 256
3357 CCAGAGGAATGCTGCTAATGTAACCAACGCGCAGCACTAAACATTAGT 3406
257 ....ThrGlyAspMetSerGlyAlaIleLysAlaIleThrLysThr 271
3407 ACAAAAACAGGGGATTAAGGTGA.....ATTGATTAACCTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSer 288
3448 CGGTAATGTAATATACAGCGAGCGCAAT..ACAATTAGGTAGTA 3494
288 snAsnThr.....Ser 291
3495 ATATCACTGTCAGATGTAACAGTAACAGCGATGCGAGCCTTGACA 3544
292 ThrThrAlaGlyAlaIleTyrValLys..... 301
3545 ACTACAGCAGGCTAACCATTAAGTCGACACAAGCAATGCAATATTAC 3554
302 .....LysLeuGluLeuAlaSerGlyL 310
3595 AACCAAAACAGGTGATCAACGGTAAGTGAATTCAGCTCCGCTGTG 3644
310 eutThrLeuPheSerArgAsnSerValAsnGlyThrAlaProLysGly 326
3645 TAAACACTGTTGCAACTGAGCACTGCTGTAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyLys..... 337
3695 AACCACTGCTACTACTCGCGATAGCGGTAAATTAACTCCACAGTAGG 3744
338 .....SerLeuSerAlaAspSerGly 345
3745 TTCTACATTAATGGAATACTGTAACCACTCAAGCAATCAGCG 3794
345 spIle.....ValPheLysGlyAsnThrValThrSerThrThrProGly 359
3795 ATATTGAAGTACAAATTCGTGTAATACAGTAATGTTCAC..... 3835

```

```

360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
3836 GCAAGCAGCTGTGATTTAACTATTGAAATAGTCAAAAGTGAAGCGAA 3885
376 uArgSerAlaAlaGlyArgAlaIleTyrPheTyrAspProIleThrThc 393
3886 AAATGGAGCTGCA..... 3898
393 LysSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898
410 AspSerAlaLeuGlnTyrThrGlyAsnIleIlePheThrGlyLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuL 443
3899 .....ACCTTAACGTGCTGAATCAGCAATTAACCCCAACAGGCT 3941
443 InProValThrLeuSerGlyLys.....ThrLeuSerLeuLysHs... 456
3942 CTGACATTACCTCAACCAATGTCAGACAACTCTTACAGCCACAGTAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATGCCAGAAATTAATGCTGCTAATGATGACGTTAAATCCACAG 4041
464 AlapherThrGlnAlaAspSerArgLeuGlnMetAspValGlyThr 479
4042 CACTTAACTACACAGGCGATTCAAAGATTACGACACAGAGTGTAACCT 4091
480 ..ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
4092 TAACAATCAATGCCAAAGATGCCAAATTAGATGCTGCTCATCAGCTGAC 4141
491 AsnLeuValIleAsnLysSerSerIleAspGlyAlaLysLysAlaLys 507
4142 CGCACAGTAAATGCACTTAACGCAAGTGGCTGTGTAACGTG..... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThr 524
4187 ....ACTGCCAAACCTCAAGCAGCGTGAATATCACCGGGGATTTAAACA 4232
524 euleuAspProThrGlyThrPheTyrGluAsn 534
4233 CAATTAATGGTTAAATATCATTTTCGCAAAAT 4264
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-302-832-8
seq_documentation_block:
; Sequence 8, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832

```

```
Filing Date: 16-SEP-1994
Classification: 435
Prior Application Data:
  Application Number: GB 9205704.1
  Filing Date: 16-MAR-1992
  Prior Application Data:
    Application Number: US pce/us93/02166
    Filing Date: 16-MAR-1993
  Attorney/Agent Information:
    Name: Belkisser, Jerry W
    Registration Number: 22,651
    Reference/Docket Number: 1038-404
    Telephone: (703) 415-0810
    Telefax: (703) 415-0810
  Information for Seq ID No: 8:
    Sequence Characteristics:
      Length: 4702 base pairs
      Type: nucleic acid
      Strandness: single
      Topology: linear
      Molecule Type: DNA (genomic)
US-08-302-832-8

Alignment Scores:
  Quality: 270.00      Length: 644
  Ratio: 0.841         Gaps: 29
  Percent Similarity: 49.845   Percent Identity: 22.671

Alignment block:
US-09-428-122-2 x US-08-302-832-8

Align seg 1/1 to: US-08-302-832-8 from: 1 to: 4702

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnPheserVa 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2537 CTAAACATCAGCGACCTTACCAACACGCTACCGCCACATTATAT 2586
43 IArgGluSerGlnGluAspAlaGlyThrThrLeuPheLysGly... 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2587 AAAACAAGAGTGTAAACCTCAAGCGGATATATCAATTAAGTGT 2636
59 . AsnValThrLeuGluAsnThrLysGlyAspLeuThrPheThr... 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2637 TAAATATCACT...ACTAAGCCTCAGCAGCTCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr... 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2684 GGAATATTAACTAACGAAAAGGCGACTTAACAATCAAGATATTAAGC 2733
88 .....GlyAsn.....GlyAsn.....GlyAsn 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2734 CGAGCGCAATCCAAATTTGGCGCAATATCTCAGAAAAGAGCGCATC 2783
92 eLeuLeuPheGln.....ThrVala 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2784 TCACAAATTTCTCTGATAAGTAATATTAACATCAGATACAAATCAAA 2833
100 AlGlyThrValAlaGlyAlaValaLysSerSerValaLysPlys 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2834 GCAGGGGTGAAGGGCGCTTCTGATCAAGTGAGCGCAAAAATGCTAA 2883
116 rThrThr.....PheI 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2884 CTTAACTATTCAACCAAGACTTAAATTTGGCAGGAGACTAAATTT 2933
120 LeuLysSerSerLeuSerPheLeuLysSerProGlySerSerLeuThr 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2934 CAGGCTTTATTAAGCAGAAATTAACGTAAATAATGCACTGATTAACT 2983
137 ThrGlyLysGlyAlaValaSerCysSerThrGlySerLeuSerLeuThr 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
2984 ATTGGC.....AATGCTAGCGGTGTAATGCTATGCTAATAAA 3021
153 S.....A 154
3022 AGTACCTTTGACACAGCTTAAGATTCAAAATCTGCACGACGGCACA 3071
154 snValSerLeuLeuPheSerLysAsnPheserThrAspAsnGlyGlyAla 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3072 ATTAACACTA.....AATAGCAGTGAAGAAAGCTTAATGTTACTAGC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSerAla.... 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3116 ..AATGCTGTAATGATACAGCACCGGTTTAACCATTTCCGCAAAAGA 3162
186 ....LeuPheSerGluAsnThrSerSerLysLys..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3163 TGTACGCTAAACATTAAGCTTACCTCCACACACAAATATATCTGTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3213 CCGCAGCAGGAAATGTACACACCAAGAGCAGCACTAATGCAACG 3262
211 GlnGlyLysValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3263 ACAGCGACGCTGCAAGTAACTGCTCAAAATGTACAAATTAAGGCAC.. 3310
227 AlIlePheThrGluAlaSerValThrIleSerAsnAlaLysValSerP 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3311 ....ATTACCTCGCAAAATGTACAGTGACAGACAGCAAGAAATCTGTA 3356
244 helIleAspAsnLysValThrGlyAlaSerSerThr..... 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3357 CCACAGAGAAATGCTGCTATTATGACACCGGACAGCAATTAATTA 3406
257 .....ThrGlyAspMetSerGlyAlaIleCysAlaThrLysThrSe 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3407 ACAAAAACAGGAGATTATTAAGTGA.....ATTGAATCAACTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerA 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3448 CGGTAAATGTAATATATACGCGAGCGGCAT..ACACTTAAGTAACTA 3494
288 snAsnThr.....Ser 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3495 ATATCACTGCTCAAGTATGACAGTACAGCGGATGCGAGCCTTGACA 3544
292 ThrThrAlaGlyAlaIleThrValLys..... 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3545 ACTACACAGCGCTCAACCATTTAGTCCGACACAGCAATGCANAATATAC 3594
302 .....LysLeuGluLeuAlaSerGlyLys 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3595 AACCAAAACAGGTATATCAACGTAAGTGAATCCAGCTCGGCTGTC 3644
310 eThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3645 TAACTCTTTGCACTGAGCACTCTGCTGATGATATTTAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyLeu..... 337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3695 AACACTTACTATTAACGCGATAGCGGTAAATTAACCTCCACACTAG 3744
338 .....SerLeuSerAlaLysSerGlyVal 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3745 TTCACAAATTAATGAGACTAATAGTAAACCACTCAGCAATCAGCGG 3794
345 spIle.....ValPheLeuGluLysThrValThrSerThrProGly 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3795 ATATGAGGTACAAATTTCTGGTAAATACAGTAAATGTTAA..... 3835
360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3836 GCAAGCACTGTGATTAATTAATGGAATATGTCACAAAGTGAACGCA 3885
```

```

376 uArgSerAlaGlyArgAlaIleTyrPheTyrAspProIleThr 393
3886 AAATGGAGCTGCA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898
410 AspSerAlaLeuGlnTyrThrGlyAsnIleIlePheThrGlyLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuG 443
3899 .....ACCTTAAGCTGCTGATACAGCAATTACCAACCAACAGAGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
3942 CTACGATTACCTCAAGCAATGTCAGACAACTTACAGCAAGATAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATCCAGGAAACATTAAATGCTGCTAATGTCAGTTAAATACCAAG 4041
464 .AlaPheThrGlnGlnAlaAspSerArgLeuGluMetAspValGlyThr. 479
4042 CACTTACTACTACAGGAGATTCAAGATTACAGCAACAGTGTACTCT 4091
480 .ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
4092 TAACAATCAATGCMAAAGATGCCAATTAGATGCTGCATCAGTGAGC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysAlaLysI 507
4142 CGCACACTAGTAATGCACTACAGCAAGTGGCTCTGTACGCG..... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThr 524
4187 ....ACTGCAAAACCTCAAGCAGCGGAAATATCACCAGGAGATTAAAG 4232
524 euLeuAspProThrGlyThrPheTyrGluAsn 534
4233 CAATAAATGGGTAAATATCATTCGGAATAAT 4264

seq_name: /cgn2_6/ptodata/2/lna/5C_COMB.seq:us-08-530-198-8

seq_documentation_block:
; Sequence 8, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-8

alignment_scores:
Quality: 270.00 Length: 644
Ratio: 0.841 Gaps: 29
Percent Similarity: 49.845 Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-530-198-8 ..
Align seg 1/1 to: US-08-530-198-8 from: 1 to: 4702

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheserVa 43
2537 CTAAACATCAGCGGACCTTTTCCCAACAACGGTACCGCCACATTAATAT 2586
43 LArgGluSerGlnGlnAlaSpAlaGlyThrThrTyrLeuPheLysGly... 58
2587 AAACACAGAGGTGTAAATCTCAAGCGCATATTATCAATTAAGTGGT 2636
59 .AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLys 74
2637 TAAATATCACT...ACTAACGCTCAGGCACATCAAAAACATTAATTAC 2683
75 SerCysPheAsnAspThrLysGlyAspLeuThrPheThr..... 87
2684 GGAATATATCACTACGAAAGAGCGCATTAACATCAACATATTAAAGC 2733
88 .....GlyAsn.....GlyAsn 92
2734 CGACGCCGAATCCAAATTTGGCGGCAATATCTCACAAAAGAGCAATC 2783
92 erLeuLeuPheGln.....ThrValAsp 99
2784 TCACAATTTCTTCTGATTAAGTAATATTACCAATCAGATACAAATCAA 2833
100 AlaGlyThrValAlaGlyAlaValaValaAsnSerSerValaLysPhe 116
2834 GCGGGGTTGAAGGGGGGCTTCTGATTCAATGAGGAGAAAATGCTAA 2883
116 rThrThr.....PheI 120
2884 CCAACTATATCAACCAAGAGTTAAATTGCGAGAGACCTAATATT 2933
120 LeglyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
2934 CAGGCTTTAATTAAGCAGAAATTACAGCTAAAATAATGCGAGTATTACT 2983
137 ThrGlyLysGlyAlaValaValaSerCysSerThrGlySerLeuSerLeu 153
2984 ATTGGC.....AATGCTAGCGGTGTAATGCTGATGCTAAAAA 3021
153 S.....A 154
3022 AGTACATTTGACAAGGTAAAGATTCAAAAATCTCGACTGACGGTCA 3071
154 snValSerLeuLeuPheSerLysAsnPheserThrAspAsnGlyAla 170

```

```

3072 ATGTACACTA.....AATACGACGATGAAAACCTAATGCTACTAC 3115
171 IleThrAlaLysThrLeuSerLeuThrglyThrMetSerAla..... 185
3116 ...AATGCTGTAAATGATTAACAGACCGGTTTAACCATTTCCCAAAAGA 3162
186 ...LeuPheSerGluAsnThrSerSerLysLys..... 195
3163 TGTACAGGTAAACAATACCTTACTCTCCACAGACATTAATATCTCTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
3213 CCGCAGCAGGAAATGTAAACACCAAGAGCAGACACTATTCATTCGACCC 3262
211 GlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
3263 ACAGGCGAGCGTGGAGTACTGCTCAAAATGTACAAATTAAGGCAAC... 3310
227 aIlePheThrGluAlaSerValThrIleSerAsnAsnAlaValSerP 244
3311 ...ATTACTCGCAAAATGTACAGTGCAGACAGCAAGAAAATCTTGTA 3356
244 belleAspAsnLysValThrGlyAlaSerSerThr..... 256
3357 CCACAGAGATGCTGCTATTAAATGCACAGCGGACACAGTAACATTAGT 3406
257 .....ThrglyAspMetSerGlyGlyAlaIleCysAlaTyrIleThrSe 271
3407 ACAAAAACAGCGGATATTAAGGTGA.....ATTGAATCAACCTTC 3447
271 rThrAspThrLysValThrLeuThrglyAsnGlnMetLeuLeuPheSera 288
3448 CGGTAATGTAAATATACACGCGCGCAAT...ACACTTAAGTAAAGA 3494
288 snAsnThr.....Ser 291
3495 ATATCACTGTCAGATGTAAACAGTAACAGCGGATGCAGAGCCTTGACA 3544
292 ThrThrAlaGlyAlaIleTyrValLys..... 301
3545 ACTACGACGCGCTCAACCATTTAGTCGACACAGCAAGCAATTAATTCAC 3594
302 .....LysLeuGluLeuAlaSerGlyLys 310
3595 AACCAAAACAGTGATATCAACGTAAGTGAATCCAGCTCCGCGCTCG 3644
310 euThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
3645 TAACTCTTGTGCACTGAGCAGCACTCTGCTGAGGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyGluLeu..... 337
3695 AACACTGTACTATTAATGCGGATAGCGGTAAATTAACCTCCACAGTAGG 3744
338 .....SerLeuSerAlaAspSerGlyAla 345
3745 TTCTACATTAATGGACTAATAGTATGTAACCACTCAAGCATTCAGGG 3794
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrThrProGly 359
3795 AATTTGAGGTACAAATTTCTGTAATACAGTAATGTATTA..... 3835
360 ThrAsnArgSerSerLeuAspLeuGlyThrSerAlaLysMetThrAlaIle 376
3836 GCAAGCAGCTGGTGAATTAACTATGGAAATGTGCAAAAGTTAAACGCG 3885
376 uArgSerAlaAlaGlyArgAlaIleTyrPheTyrAspProIleThrThrG 393
3886 AATGTGAGCTGCA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898

```

```

410 AspSerAlaLeuGlnTyrThrGlyAsnIleIlePheThrGlyGluLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuG 443
3899 .....ACCTTAAGTGTGATCAGGCAATTAACCAACCAACAGGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
3942 CTACGATTAAGTCAACATGCTCAGACACTCTTACAGGCANAGGCTAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATCGCAGAAACATTAATGCTGCTAATGTGCGTTAAATCCACAG 4041
464 .AlaPheThrGlnGluAlaAspSerArgLeuGluMetAspValGlyThr 479
4042 CACTTACTACTACAGGGGATTCAAAGATTACGCAACGACAGGTGACT 4091
480 ...ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
4092 TAACATCAATGCCAAAGATGCCAAATTAGATGCTGCTGATAGGTGAC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysLysAlaLysTl 507
4142 CGCACAGTAAATGCAATCACTACGCAAGTGCGCTGCTGTAAGTGC... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThr 524
4187 ...ACTGCAAAACCTCAGACGCGTGAATATCACCGGAGATTAAACA 4232
524 euLeuAspProThrGlyThrPheTyrGluAsn 534
4233 CAATAAATGGTTAAATATCACTTCGCAAAAT 4264

seq.name: /cgn2_6/prodata/2/ina/SC_COMB.seq:US-08-469-880-8

seq.documentat.ion_block:
? Sequence 8, Application US/08469880
? Patent No. 5876733
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J.
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? TITLE OF INVENTION: of No. 5876733-typeable Haemophilus
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Mattare, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? STREET: Bldg. 1
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/469,880
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9205704.1
? FILING DATE: 16-MAR-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? PRIOR APPLICATION DATA: US 08/302,832
? APPLICATION NUMBER:
? FILING DATE: 16-SEP-1994

```



```
ATTORNEY/AGENT INFORMATION:
NAME: Betkresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4702 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-8

alignment_scores:
Quality: 270.00      Length: 644
Ratio: 0.841        Gaps: 29
Percent Similarity: 49.845      Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-469-880-8 ..

Align seg 1/1 to: US-08-469-880-8 from: 1 to: 4702

27 LeuaspSerSerAlaSerPheaspGlyAsnLysasnGlyasnPheserVa 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2537 CTAACATCTACCGGACCTTACCACACAGCATCCGCCCAACATTATAT 2586
43 LAGgluSerGlnLuspaLagIyThrThyLeuPhelysGly... 58
.....:.....:.....:.....:.....:.....:.....:
2587 AAAACAAGGAGGTGTAATACTCCAGGCGATATTATCAATPAAGGTGT 2636
59 .AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThLys 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2637 TAATATACCT...ACTACGCGCTCAGCAGCTCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr..... 87
:::.....:|||||:|||||:|||||:|||||:|||||:
2684 GGAATAATACTAACAAGAAAGCGACTTAACATCAAGATATTAAAGC 2733
88 .....:.....:GlyAsn.....:GlyAsn 92
|||||:|||||:|||||:|||||:|||||:|||||:
2734 CGACGCCGAATCCAAATTCGCGCATATCTACAAAGAAAGGCAATC 2783
92 erLeuLeuPheGln.....:.....:ThrValAsp 99
:::.....:|||||:|||||:|||||:|||||:
2784 TCACAATTCTCTGATAAAGTAATATTACCAATCAGATAACATCAAA 2833
100 AlagIyThrValAlaGlyAlaAlaValAsnSerSerValAlasplys 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2834 GCAGGCGCTTAAGGGGGCTTCGTGATTCAGGTAGGAGAAATGCTAA 2883
116 rThrThr.....:.....:PheI 120
:::|||||:|||||:|||||:|||||:|||||:
2884 CTAACATTCCTCAACCAAGAGTTAAATTGGCAGACCTTAATATT 2933
120 leGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2934 CAGGCTTTAATAAGCAAGAAATTACAGCTAAATAATGCACTGATT 2983
137 ThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThr 153
|||:|||||:|||||:|||||:|||||:|||||:|||||:
2984 ATTGGC.....:AAATGCTAGCGGTGTATGCTGATGCTAAAA 3021
153 S.....:.....:A 154
|||:|||||:|||||:|||||:|||||:|||||:|||||:
3022 AGTAGCTTTGACAGAGTTAAAGATTCAAAATCTGACTGACGCTGACA 3071
154 snValSerLeuLeuPheSerLysAsnPheserThrAspAsnGlyAla 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
3072 ANGTACACTA.....AATAGCGAAGTGAAGAAAGCTTAATGCTAGTAC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThreSerAla..... 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3116 ...AATGCTGTGTAATGATAACAGACCGGTTTAACCATTTCCGCAAGA 3162
186 ...LeuPheSerGluAsnThrSerSerLys.....:..... 195
:::|||||:|||||:|||||:|||||:|||||:|||||:
3163 TGTAAAGGTAAACAAATAGCTTACCTCCCAAGAACAAATTAATCTCTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3213 CCGCAGCAGCAAAATGTAAACACCAAGAAAGCAGCACTTCATAGCAACC 3262
211 GlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAla 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3263 ACAGCGCAGGTGGAAGTAACTGCTCAAAATGTATCAATTAAGGCANAC 3310
227 aIlePheThrGluAlaSerValThrIleSerAsnAlaLysValSerP 244
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3311 ...ATTACCTCGCAAAATGTAAACAGTACAGCACGACAAAGAAATCTT 3356
244 heIleAspAsnLysValThrGlyAlaSerSerThr.....:..... 256
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3357 CCACAGAGAAATGCTCTCATTAATGCAACCGCGGCACTTAACATTACT 3406
257 .....ThrGlyAspMetSerGlyAlaIleCysAlaTyLysThre 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3407 ACAAAAACAGGGGATTTAAAGTGA.....ATTGAATCAACCTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuPheSerA 288
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3448 CGGTAAATGTAATATTACAGCGAGCGCAAT...ACACTTAAGGTAA 3494
288 snAsnThr.....:.....:Ser 291
|||:|||||:|||||:|||||:|||||:|||||:|||||:
3495 ATATCACTGCTCAAGATGTAAAGTAAACGAGGATGACGAGCTTGACA 3544
292 ThrThrAlaGlyAlaIleTyValLys.....:..... 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3545 ACTACGACGAGCTCAACATTAAGTGGCAGACAGCAATGCAAAATTAC 3594
302 .....:.....:LysLeuGluLeuAlaSerGlyL 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3595 AACCAAAACAGTGATATCAAGGTAAAGTTGAATCAGCTCGGCTCG 3644
310 eutThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3645 TAACACTTGTTCGAACTGAGCAACCTTGCCTGAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyLysLeu.....:..... 337
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3695 AACACTGTACTATTCTCGCGATAGCGTAATTAATTAACCTCCAGCTAG 3744
338 .....:.....:SerLeuSerAlaAspSerGlyA 345
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3745 TTCTACATTAATGAGCTAATAGTGTACACCTCAACCAATCAGCGCG 3794
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrThrProGly 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3795 ATATTAAAGTAACTTCTGTGAATACAGTAATGTTCA.....:..... 3835
360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3836 GCAGAGACGTGTGATTACTATTGGAATAGTGCAGAAAGTTGAAGCGAA 3885
376 uArgSerAlaIleGlyArgAlaIleTyThrPheTyraspProIleThr 393
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3886 AATGAGCTGCA.....:.....: 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGlnThrProAla 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3898 .....:.....: 3898
```

```

410 AsperalaleuGlnTyrThrGlyAsnIleIlePheThrGlyIuysIe 426
3898 ..... 3898
426 userGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuG 443
3899 .....ACCTTAAGCTGATCAAGCAATTAACCAACCAAGGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuYsHs... 456
3942 CTACAGTACCCTCAAGCAATGCGACACACTTACAGCCAGAGATAC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATCCAGAAACATTAATGCTGCTATGATGAGTTAAATACACAG 4041
464 .AlaPheThrGlnGlnAlaAspSerArgLeuGluMetAspValGlyThr 479
4042 CACTTAACCTACTACAGGGGATTAAGATTAACGCAACGAGGTACCT 4091
480 ..ThrLeuGluProIleAsp.....ThrSerThrIleAsn 490
4092 TAAACAATCAATGCAAAAGATGCCAATTAAGTGTGCTGATGAGTAC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysAlaLysI 507
4142 CGCAGCTAGTAAATGCAACTACGCAAGTGTGCTGTAACGTG..... 4186
507 eglAThrlYsAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThr 524
4187 ...ACTGCAAAACCTCAAGCAGCGTGAATATACCGGGGANTTAACA 4232
524 euleuAspProThrGlyThrPheYrGluAsn 534
4233 CAATAATGGTAAATATCATTTGCGAAAT 4264

seq.name: /cgn2_6/ptodata/2/ina/50_COMB.seq:US-08-728-470-8

seq.documentation_block:
: Sequence 8, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4702 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-8

Alignment_scores:
: Quality: 270.00 Length: 644
: Ratio: 0.841 Gaps: 29
: Percent Similarity: 49.845 Percent Identity: 22.671

Alignment_block:
US-09-428-122-2 x US-08-728-470-8 ..
Align seq 1/1 to: US-08-728-470-8 from: 1 to: 4702

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerVa 43
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2537 CTAAACATCAACGCGGCGCTTACCAACACGATACCGCCACATTAATAT 2586
43 IArgGluSerGlnGluAspAlaGlyThrThrYrLeuPheLysGly.... 58
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2557 AAAACAAGAGGTGTAACCAACGCGATATTATCAATAAAGGTGCTT 2636
59 ..AsnValThrLeuGluAsnIleProGlyThrAlaIleThrLys 74
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2637 TAAATATCACT...ACTAACGCTCAGCAGCTCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr..... 87
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2684 GGAATATTAACTAACGAAAGGCGACTTAACATCAAGATATTAAAGC 2733
88 .....GlyAsn.....GlyAsn 92
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2734 CGAGCGCAATCCAAATGGCGCATATTCACAAAGAAAGGCGATC 2783
92 eulLeuLeuPheGln.....ThrValAsp 99
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2784 TCACAAATTCCTGCTAAGTAATATTACCAATCAGATTAACATCAAA 2833
100 AlaGlyThrValAlaGlyAlaValAsnSerSerValValAspLysSe 116
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2834 GCAGGCGTGAAGGGGCGTTCGATTCAAGTGAAGCAAGAAATGCTAA 2883
116 rThrThr.....PheI 120
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2884 CTAACTATTCAACCAAGATTAATGCGAAGAGACCTAAATATT 2933
120 IeGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2934 CAGGCTTATTAAGCAAGAAATTACAGCTAAATATGCAAGATTTAACT 2993
137 ThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrly 153
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2984 ATGGC.....AATGCTAGCGGTGTGAATGCTGATGCTAAAAA 3021
153 S.....A 154
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
3022 AGGACCTTTGCAAGGTAAAGTTCAAAATCTGCACTGAGCGGTACA 3071
154 snValSerLeuLeuPheSerLysAsnPheSerThrAspAsnGlyAla 170
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:

```

```

3072 ATGTACACTA.....ATAAGCAAGTGAAGAACGCTAATGCTAGTAGC 3115
171 lIethrAlaLysThrLeuSerLeuThrGlyThrMetSerAla.... 185
      ::::: ::::: ::::: :::::
3116 ...AATGCTGAATGATACAGCACCGCTTAACCATTTCCGCAAGA 3162
186 ...LeuHeSerGluAsnThrSerSerLysLys..... 195
      ::::: ::::: ::::: :::::
3163 TGTACAGGTAACATTAACGTTACTCCACAGACAAATAATATCTCTG 3212
196 .....GlyGlyAlaLeuIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
      ::::: ::::: ::::: :::::
3213 CCGCAGCAGGAATGTATACACCAAGAGACACATATCATGATCAGC 3262
211 GlnGlyGluValSerPheSerAspAsnThrSerAspSerGlyAlaI 227
      ::::: ::::: ::::: :::::
3263 ACAGGCACCGTGAAGTAACTGCTCAAAATGTACAAATTAAGGCAAC 3310
227 aIllePheThrGluAlaSerValThrIleSerAspAsnAlaLysValSerP 244
      ::::: ::::: ::::: :::::
3311 ...ATTACCTCGCAAAATGTATACAGTACAGCAACAGAAATCTTGT 3356
244 heIleAspAsnLysValThrGlyAlaSerSerSerThr..... 256
      ::::: ::::: ::::: :::::
3357 CCACAGAGAAAGTGTCTATTAATGCAACCGCAGCAAGTAACATTA 3406
257 .....ThrGlyAspMetSerGlyGlyAlaIleGlyAlaIleGlyThrSe 271
      ::::: ::::: ::::: :::::
3407 ACAAAAACAGGGGATTAAGGTGA.....ATTGAATCAACTTC 3447
271 tThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSer 288
      ::::: ::::: ::::: :::::
3448 CGGTAATGTAAATATTACAGGAGCGGCAT...ACACTTAAGTAAGTA 3494
288 snaSnThr.....Ser 291
      ::::: ::::: ::::: :::::
3495 ATATCACTGTCAGAGATGTAAAGTACAGGAGGATGAGAGCCTTGACA 3544
292 ThrThrAlaGlyGlyAlaIleTyValLys..... 301
      ::::: ::::: ::::: :::::
3545 ACTACAGAGAGCTCAACCATTAAGTGCACACAGGCAATGCAATATTAC 3594
302 .....LysLeuGluLeuAlaSerGlyLys 310
      ::::: ::::: ::::: :::::
3595 AACCAAAACAGGTGATATCAACGGTAAGTGAATCCAGCTCCGGCTGTG 3644
310 euThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
      ::::: ::::: ::::: :::::
3645 TAAACTGTGTGCAACTGGAGCACTCTTGCTGTAAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyGluLeu..... 337
      ::::: ::::: ::::: :::::
3695 AACACTGTACTATTACTGCGGATAGCGGTAAATTAACTTCACACATGAG 3744
338 .....SerLeuSerAlaAspSerGly 345
      ::::: ::::: ::::: :::::
3745 TTCTACATTAATGAGACTAATAGTGAACCAACCAAGCAATCAGAGCG 3794
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrThrProGly 359
      ::::: ::::: ::::: :::::
3795 ATATTGAAGTACATTTCTGTATATACAGTAATGTTACA..... 3835
360 ThrAsnArgSerSerLeuAspLeuGlyThrSerAlaLysMetThrAlaLe 376
      ::::: ::::: ::::: :::::
3836 GCAGCACTGCTGATTTAATGTAATGTAATGTAATGTAATGTAATGTA 3885
376 uArgSerAlaAlaGlyAlaIleTyPheTyAspProIleThrG 393
      ::::: ::::: ::::: :::::
3886 AAATGGAGCTCA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
      ::::: ::::: ::::: :::::
3898 ..... 3898

```

```

410 AspSerAlaLeuGlnIleTyThrGlyAsnIleIlePheThrGlyGlyLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuG 443
      ::::: ::::: ::::: :::::
3899 .....ACCTTAAGTGTGATACAGCAAAATTAACCAACCAACAGGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
      ::::: ::::: ::::: :::::
3942 CTACACTTACTCTACAGCAATGTGTACAGCACTCTTACAGGCAAGATAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATCCAGCAAAATTAATGCTGTATATGTCAGTTAATACCAAGC 4041
464 .AlaPheThrGlnGlnAlaAspSerArgLeuGluMetAspValGlyThr. 479
      ::::: ::::: ::::: :::::
4042 CACTTAACTACTACAGGGGATTTCAAGATTAACGCAACCAAGTGTACT 4091
480 ..ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
      ::::: ::::: ::::: :::::
4092 TAAACATCAATGCAAAAGATGCCAAATTAAGATGTGCTGCTCAGTCAG 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysLysAlaLysII 507
      ::::: ::::: ::::: :::::
4142 CGCAGCTAGTAATGCACTTAACGCAAGTGTGCTGTGTAACGTG..... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrL 524
      ::::: ::::: ::::: :::::
4187 ...ACAGCAAAACCTCAAGCAGCGGATATACCGCGGATTTAACA 4232
524 euLeuAspProThrGlyThrPheTyGluAsn 534
4233 CAATTAATGGTTAATATCATTCGCAAAAT 4264
seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:us-08-617-697-8
seq_documentation_block:
; Sequence 8, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651

```



```

1524 TCAACGGCGCTGTGAGATCAACGACTCTGCGCGCTCGCCACGAC 1573
      ::::::::::::::::::::
241 ..... LysValSerPheIleAspAsnLysValThrGlyAla..... 252
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1574 GCGGCAAGATGCGCAGCGTCAACCTGGGCACTTGGGGCGCGCACGAT 1633
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
253 ..... SerSerSerThrThrGlyAspMetSerGly..... 262
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1624 CGAGTCAGCGCTCTGAGACGCTCAACCTGTGGGCGACGGCAGCTCGC 1673
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
263 ..... GlyAlaIleCysAlaThrLysThrSerThrAspThr 274
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1674 TCGGCATCGCGCGCGCTCTGACGCC..... 1702
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
275 LysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAspThr 291
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1703 ..... ACCGCGACCGCCACACCTGACCTGACGTCATGCTGAC 1746
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
291 rThrThrAlaGlyGlyAlaIleThrValLysLysLeuGluLeuAlaSer 308
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1747 GAGGAC..... GCGCGATCGAGACTGGAAAGCGCTGCTGACGATG 1790
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
308 LysLeuThrLeuPheSerArgAsnSerValAsnGlyLysThrAlaPro 324
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1791 GT..... TTCACCCACATCAACATCGTGTGACCGCCCTCT 1828
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
325 LysGly... GlyAlaIleAlaIleGlyAspSerGlyLysLeuSerLeu 340
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1829 TCGACGATGCGCAGCTGTGGCGCGCGACGACGACCCCTCAACATCTC 1878
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
340 rAlaAspSerGlyAspIleVal..... PheLeuG 350
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1879 GGGCGAGCTCGCTCAGATCAGTCTGACACCCGCTGCCCTGACGG 1928
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
350 LysAsnThrValThrSerThr..... 356
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1929 GCATTCAGGTGACCAACAGCGTGTGCGACCTCGCGCCGCACTGGC 1978
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
357 ..... ThrProGlyThrAsnArgSerSerIleAspLeuG1 368
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1979 ACCGCTGTGCTTCACGGCGGCGCTGGCGCTGACTGATCTCTGG 2028
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
368 YThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyAlaIleT 385
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2029 CGCGCAGACCAAG..... GCGATGTCATGGGCGCGCGC..... 2062
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
385 YrPheThrAspProIleThrThrGlySerSerThrThrVal..... 398
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2063 ..... GACGACACCGTCAACGCTGAGCTGGGAGCCCTGGCCTGCT 2104
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
399 ..... ThrAspValLeuLysValAsn..... 405
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2105 GGTTCGGTCAACGGCGGCGGACGACGACGCTGTCTGTGGCCAAAGTCAA 2154
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
406 ..... GluThr..... 407
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2155 CGGTTCGTGCTCAGCGCTGACCGCGCTTGGCGCTTCGAACCTCTCC 2204
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
408 ... ProLysAspSerAlaLeuGlnThrGlyAsnIleLe... PheThr 422
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2205 GCGTCCGTGGCGCGCTCAAGGCTCGCAACACGCCAAGCGCTTCAG 2254
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
423 GlyGluLysLeuSerGluThrGluAlaAla..... 432
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2255 GCTCTCACTGGCGGCGGCGGTGCGACGACCTTCACCAACGTTGC 2304
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
432 ..... 432
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2305 GGTGAATGTGCGCTGACCGTCTTGGCGGCTCGACCGGTACGACGACG 2354
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
433 ..... AspSerLysAsnLeuThr 439
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

2355 TGACCCCTGGCCACGCCACGGCAGCTCGAGCTGTCAACCTGAC... 2401
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
440 LysLeuLeuGlnProValThrLeuSerGlyLysLeuSerLeu..... 454
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2402 ... CTGTGCTCTGGCGCTCTGCGCGCTGTGAGGTTGGTGGTCTGG 2448
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
455 ..... LysHisG 457
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2449 GCTGACGAGTGAATGCGCGCCAGCAGACCAACGACCGCTCAG 2498
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
457 LysValThrLeuGlnThrGluAla..... PheThr 466
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2499 TCGACAGCTGACGCTGCAAGCCAGCTCGCCAGATGATGATGAGCAG 2548
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
467 GlnGlnAlaAspSerArgLeuGlnMetAspValAlaThrThrLeuIuP 483
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2549 GGCACGCGCGTGTGACCTGACACACCGCGACACGCGCTGACAG 2598
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
483 rAlaAspThrSerThrIle..... 489
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2599 CTTCGACGCGCAGCGCGCTCAGCGGACGCTCGGCTGTGACCTTGCT 2648
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
490 ..... AsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2649 CGGCCAACACACGCTGGGTGAAGTGTACGATCCGCGCGCGCTGSC 2698
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
505 AlalysIleGluThrLysAlaThrSerLysAsnLeuThrLeu..... 518
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2699 GCGACCTGCTGACCGGCTGGCGCACCGCAATGACACATCATCGGTGG 2748
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
519 SerGlyThrIleThrLeuLeuAspProThrGlyThrPheThrGluAsnH 535
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2749 CGCTGGCTACACCTGTGTACACCGCGCGGTAC.....G 2786
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
535 LysSerLeuArgAsnProGlnSerThrAspIleLeuGluLeuLysAlaSer 551
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2787 ACACCTTCACGGGTGGCGGCGCGGATCTTGTGATATCAACGCTATC 2836
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
552 GlyThrValThrSerThrAlaValThrProAspProIleMetGlyLys 568
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2837 GGCACCTTCGACCGCTTTCGTGACGATCACCGCCGCTGTGCGCACAA 2886
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
568 sPheHisThrGlyThrGlnGlyThrThrProIleValThrGlyThr. 584
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2887 GCTGACCTGCTGGCGATCTCGACGACGCGGCTATGCTGACGCGGCT 2936
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
585 ... GlyAlaSerThrThr 589
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2937 TCGGCGCTGCGTCAAC 2953
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-614-377A-6
seq_documentation_block:
; Sequence 6, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smith, John
; APPLICANT: Bingle, Wade H.
; TITLE OF INVENTION: NO. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HEREROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

385 yrhethyrasprothlethrthrclysersethrthrvall..... 398
      |||
2063 .....SAGCAGACCGCTCAGCTCGGCGGCGCTGCT 2104
399 .....ThrAspValLeuLysValAsn..... 405
2105 GATTGCTGACAGCGCGGCGGAGCGACCGAGCTTCTGTGCGCCAGCTCAA 2154
406 .....GluThr..... 407
2155 CGGTGCTGCTCAGCGCTGACCGGCTTGGCGGCTTCGAAACCCCTCC 2204
408 ..ProLAspSerAlaLeuGlnThrthrclyserlelle...PheThr 422
      |||
2205 GCGTCCCTGCGCGCGGCTCAGGCTCGACACCGCCAGCGCTTCACG 2254
423 GlyGluLysLeuSerGluThrGluAlaAla..... 432
2255 GCTCTGACACTGGCGCGGCGGCTGCGAGCCTTCACCAACGTTGC 2304
432 ..... 432
2305 GGTGAATGTGGCTGACCGCTTTCGGCGGCTCCGACCGGTACAGACCG 2354
433 .....AspSerLysAsnLeuThrSer 439
2355 TGACCTGTGCGCACACCGCGGCGGCTGAGCGTGTTCACCTGAC... 2401
440 LysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLeu.... 454
      |||
2402 ..CTGTGCTCTGCGCGGCTCTGCGGCTGTGAGGTTGCGCTGCGG 2448
455 .....LysHisG 457
2449 CGTCGAGAGCGTGAAATCGCCGCGCACCGACACCAACAGACCGCTCAG 2498
457 LysValThrLeuGlnThrGluAla.....PheThr 466
2499 TCGACAGCGTGTGAGCTGACAGCGCACCTCGGCGAGTCGTGTGTGAG 2548
467 GlnGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluPr 483
2549 GGCACAGCGGCTGTGACCTGACCAACACCGGACACCGCTGTACAG 2598
483 oalaAspThrSerThrile..... 489
2599 CTTCGACGCGCAGCGCGCTCACCAGCGGCTCGGCTGTGACCTCGTGT 2648
490 .....AspAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
2649 CGGCGCAACACAGCGGTGGGTAAATGCTCACATCCGCGCGGCGCTGAG 2698
505 AlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeu..... 518
2699 GCGGACTCGCTTACCGGTTGCGCCAGCGCAATGACCAATCATATGCGTGG 2748
519 ..SerGlyThrIleThrLeuLeuAspProThrGlyThrPheYrGluAsn 535
2749 CGGTGCGCTGACACCGCTGTCTACACCGCGGTAG.....G 2786
535 IserLeuArgAsnProGlnIserThrAspIleLeuGluLeuLysLaser 551
2787 ACACCTTCACGGGTGGCGCGCGGATATCTCATATCACCGCTTC 2836
552 GlyThrValThrSerThrAlaValThrProAspProIleMetGlyGly 568
2837 GGCACCTCGACCGCTTGTGTCGATCACCGACCGGCTGTGCGGAA 2886
568 srPheHisYrGlyYrGlnGlyThrTrpGlyProIleValThrIpsGlyThr 584
2887 GCTCGACCTCTGCGCATCTGACGAAACGCGCTATGCTGACGCGGCT 2936

```

```

585 ..GlyAlaSerThrThr 589
2937 TGGGCGTGGCGTACC 2953

seq_name: /cgn2_6/prodata/2/1na/5A_COMB.seq:US-08-038-682-5
seq_documentation_block:
; Sequence 5, Application US/08038682
; Patent No. 5549897
;
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Mataro, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-038-682-5

alignment_scores:
      Quality: 241.50      Length: 614
      Ratio: 0.736      Gaps: 28
Percent Similarity: 53.420      Percent Identity: 22.638

Alignment_block:
US-09-428-122-2 x US-08-038-682-5 ..

Align seg 1/1 to: US-08-038-682-5 from: 1 to: 9171

20 MetIleAlaThrGluThrValLeuAspSerSerAlaSerPheAspGlyAs 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2752 ATGATTGGTAAAGCATGTACCAAAAATAACATACCTTGAAGGAGG 2801
36 nlyAsnGlyAsnPheSerValArgGluSerGlnGluAspAlaGlyThrT 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2802 TAAAGTG...AGGTTGGCTCCAGAAAGCGTACAGAA..... 2838
53 hTrYLeuPheLysGlyAsnValThrLeuGluAsnIleProGlyThrGly 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2839 .....ATCGAAGCAATGTACTATCAATAAC.....AACGCTAAC 2874
70 ThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThr.. 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


2875 GTCACTTTATCGGTTCGATTTCACACCACTCAAAACCTTTAACTAT 2924
 86PheThrGlyAsnGlyAsn 92
 2925 TAAAAAGATGTCATCATTAATAGCGCAACCTTACCGCTGAGCGCAATA 2974
 92 erLeuLeuPheGlnThrValAspAlaGlyThrValAlaGly..AlaAla 107
 2975 TTGTC.....AATATACCGCGAAATCTTAC 3000
 108 ValAsnSerSerValValAspLysSerThrThrPheIleGlyPheSerSe 124
 3001 GTTGAAGATAGCGTAATTTCACAGCTATCACAAATTCACCTTTAATGT 3050
 124 rLeuSerPheIleAlaSerProGly...SerSerIleThrGlyLysG 140
 3051 AGCGCGCTGTTTGACACAAAGCAATCAAAATTTCCATTGCCCAAG 3100
 140 lValAlaSerCys.....SerThrGlySerLeuSerLeuThr 152
 3101 GAGGCGCTCGCTTTAAGACATGATTAATCCAGAAATTTAAGCATCC 3150
 153 LysAsnValSer.....LeuLeuPheSerLysAsnPheSerThr 165
 3151 ACCAAGCTCCAGCTCCACTACCGCACTATTATAGCGCAATATAACCA 3200
 165 rAspAsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThr.....G 180
 3201 TAAAAACGGT.....GATTTAAATATTACGAAGANG 3232
 180 lYThrThrMetSerAlaLeuPheSerGluAsnThrSerSerLysGly 196
 3233 GTAGGATACGTAATGCAATGCGGCGGATGCTCCGCAAAAGAGGT 3282
 197 GlyAlaIleGlnThrSerAspAlaLeuThrIleThr..... 208
 3283 AATCTCAGCATTTCTTGCAAAATCAATATTACCAACAGATTAACAT 3332
 209GlyAsnGlnGlyLysValSerPheSerAspAsnThrSerSe 223
 3333 CAAGCAGAGGTGTGATGGGAGAAATTCGATTCCAGCGCACAAACATG 3382
 223 spSerGlyAlaAlaIlePheThrGluAlaSerValThr..... 235
 3383 CCAATCTAACCATTAACCAAGAAATGAATTAACGCAAGACTTAAT 3432
 236 lIleSer.....AsnAsnAlaLysValSerPheIleAspAsnLysValTh 250
 3433 ATTTCAAGTTTCATTAAGCAAGAGATTTACAGCTAAAGATGTAGTAT 3482
 250 rGlyAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysA 267
 3483 AACTTTGTTGTAACCAATAGTGTGATGCTACTATGCCAAAAAAAGTAA 3532
 267 lAtyLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 283
 3533 CCTTAACAGGTTAAAGATTCAAAATCTCTGCGAGCGGTCAACAAGTG 3582
 284 LeuLeuPheSer.....AsnAsnThrSerTh 292
 3583 ACATCTACACAGCAAAAGTGCAACATCCGGTACTAATAACACACATGA 3632
 292 rThrAlaGlyGlyAlaIleTyValLysLysLeuGluLeuAlaSerGlyG 309
 3633 TAGCAGTGCAAT.....AATGCCG 3652
 309 lYLeuThrLeuPheSerArgAsn...SerValAsnGlyGlyThrAlaPro 324
 3653 CCTTAACATTCATGCAAAATGTAACAGTAACACAAAT..... 3693
 325 LysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeuSerLeuSerAl 341
 3694ATTACTTCTCAAAAGAGTGAGCATCTCTGC 3725

341 aAspSerGlyAspIleValPhe...LeuGlyAsnThrValThrSerThrT 357
 3726 GACAGGTGGAAMAATTACACTAAACAGGTACAAACCATTAACCAACA 3775
 357 hrrProGly.....ThrAsnArgSerSerIleAspLeuGly..... 368
 3776 CTGTACAGGTGAGATACCGCTCAACAGGTAGTATCTGAGTGGAATT 3825
 369ThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAl 383
 3826 GAGTCCAGCTCTGGCTGTGTAACACTTACTGCAACCGAGCGCTCTGC 3875
 383 aIleTyPheTyAspProIleThrThrGlySerSerThrThrValThra 400
 3876 TGT.....ACCAATATTTCGCGCAACCGCTACTGTACTGTG 3913
 400 spValLeuLysValAsnGlnThrProAlaAspSerAlaLeuGlnTyThr 416
 3914 CAATAGCGGTGCTTAACCACTTTGGCAGGCTCTACATTAAGCAAC 3963
 417 GlyAsnIleIlePheThrGlyLysLeuSerGluThrGluAlaAlaAs 433
 3964 GAGACTGA.....ACCACTTCAAGTCA 3986
 433 pSerLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGly 450
 3987 ATCAGCGCATTCGCGGT.....ACGATTTCTGTGTG 4018
 450 lYThrLeuSerLeuLysHisGlyValThrLeuGlnThrGluAlaPheThr 466
 4019 GCACGTAAGGTTAA.....GCAACCGAAAGTTTAAC 4053
 467 GlnGluAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluPr 483
 4054 ACTCATTCATTCAAAATTAAGCAACACAGCGGCTAAGCTAAC 4103
 483 oAlaAspThrSerThrIleAsn.....AsnL 492
 4104 AAGTGCAACAGGTACAAATTTGGTGTAGCATTTCCGTAATCGGTAA 4153
 492 euValIleAsn.....IleSer 497
 4154 TTACGGCAACGCTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAAT 4203
 498 SerIleAspGlyAla.....LysLysAlaLysIleGlnTh 509
 4204 GCGACAGAGAGCTGCACCTTAACATCAATCGGCAAAATTAACATAC 4253
 509 rLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeuA 526
 4254 CGAAGCTAGTTCACACATTAATCAAGCAAGGTCAAGTAATCTTTCAG 4303
 526 spProThrGlyThrPheTyGluAsnHisSerLeuArgAsnProGlnSer 542
 4304 CTCAGAGATGTAACCTGCAGAGATTAATATGCCCAAT..... 4344
 543 TyrAspIleLeuGluLeuLysAlaSerGlyThrValThrSer 556
 4345GTGACACTAAATATCAAGGCACTTAACCTAC 4377

Fri Nov 24 13:50:08 2000

us-09-428-122-2.rni

Page 24

OM of: US-09-428-122-2 to: EST:* out_format: pfs

Date: Nov 20, 2000 3:57 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-O/cgcn2.1/USPTO.spool/US09428122/runat.14112000.163116.27900/app_query.fasta.1.999
-DB=EST -OFMT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELPE=6.000
-DELPEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09428122.@CGN1.1.2757 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-428-122-2
Query length: 928

Database sequences: 7189864

Database length: -1203564053

Search time (sec): 6294.330000

Sequence list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_gss23:CNS04ALV	171.00	304.40	5.4e-08	898	AL282028 Tetraodon nigroviridis
gb_gss23:CNS04RM2	158.50	281.26	1.1e-06	805	AL304427 Tetraodon nigroviridis
gb_est15:AU062062	140.00	246.19	9.5e-05	730	AU062062 Tetraodon nigroviridis
gb_gss21:CNS00GFX	131.50	225.33	0.0014	1006	AL0072139 Drosophila melanogaster
gb_gss12:AO875306	129.50	229.95	0.0008	520	AO875306 V122E4 mtm-3xHA/lacZ
gb_est1:AM256683	127.50	221.68	0.0022	727	AM256683 EST304820 KV2 Medicago
gb_gss9:AO645402	127.00	221.70	0.0022	673	AO645402 RRC193-DpnII-29D12.T9
gb_est18:AV403135	126.00	223.38	0.0017	500	AV403135 AV403135 Bombyx mori
gb_est18:AV403126	126.00	220.19	0.0027	650	AV403126 AV403126 Bombyx mori
gb_est1:AI16701	125.00	217.39	0.0038	613	AI16701 xylem.est.507 Poplar x
gb_est11:AI151569	124.00	217.02	0.0040	613	AI151569 GH28575.5prime GH Dros
gb_gss24:CNS05GDN	123.50	208.84	0.0115	1068	AL336164 Tetraodon nigroviridis
gb_gss9:AO640209	123.00	215.79	0.0047	579	AO640209 927P1-2P1.TV 927P1.TV
gb_est8:AI115059	121.00	207.35	0.0139	820	AI115059 ui41h04.y1 Sgano mos
gb_gss23:CNS03TE7	121.00	205.89	0.0167	918	AL289756 Tetraodon nigroviridis
gb_est7:AF031703	121.00	202.99	0.0243	1149	AF031703 AF031703 Human lymph
gb_est25:AM906712	120.50	207.20	0.0141	769	AM906712 EST342835 potato stoid
gb_gss9:AO639357	120.00	208.68	0.0117	655	AO639357 927P1-10F10.TV 927P1.T
gb_est4:BE247955	120.00	208.30	0.0123	655	BE247955 NF002D12D1F1094 Drog
gb_est15:AU061810	120.00	207.15	0.0142	716	AU061810 AU061810 Dictyostelium
gb_gss23:CNS04TIL	120.00	204.61	0.0197	871	AL306930 Tetraodon nigroviridis
gb_gss21:AI151361	119.50	206.53	0.0154	696	AI151361 GH28540.5prime GH Dros
gb_est21:AM256769	119.50	206.36	0.0157	705	AM256769 EST304906 KV2 Medicago
gb_est16:C93043	119.50	205.77	0.0170	738	C93043 C93043 Dictyostelium dis
gb_est13:BE130661	119.50	205.26	0.0182	768	BE130661 L48-86273 Ice plant La
gb_est13:BE034472	119.50	205.22	0.0182	770	BE034472 MH05E02 MH Mesembryant
gb_est13:BE130655	119.50	205.10	0.0185	777	BE130655 L48-85673 Ice plant La
gb_est13:BE131175	119.50	204.91	0.0190	789	BE131175 L48-115073 Ice plant La
gb_est13:BE033730	119.50	203.69	0.0222	867	BE033730 MF07F05 MF Mesembryant
gb_est13:BE036160	119.50	202.26	0.0267	968	BE036160 MO20D11 MO Mesembryant
gb_est13:BE035792	119.50	201.96	0.0279	995	BE035792 MO10C04 MO Mesembryant
gb_gss23:CNS04NSM	119.00	204.84	0.0191	735	AL289119 Tetraodon nigroviridis
gb_est23:AM640354	119.00	203.98	0.0225	810	AM640354 bl94c08.w1 Blackshear
gb_gss22:CNS02PLH	119.00	201.92	0.0278	921	AL208142 Tetraodon nigroviridis
gb_gss23:CNS03XRI	119.00	199.62	0.0340	781	AL265383 Tetraodon nigroviridis
gb_est14:AU004371	118.50	199.07	0.0240	1100	AU004371 AU004371 Bombyx mori
gb_est13:BE035623	118.50	199.52	0.0379	1028	BE035623 MO12D08 MO Mesembryant
gb_est13:AM055690	118.00	207.14	0.0143	529	AM055690 rs38e02.y1 Sommer pfs
gb_gss24:FR003380	118.00	205.14	0.0184	617	AL030167 Fugu rubripes GSS seq
gb_gss21:CNS00EDS	118.00	197.65	0.0482	1101	AL068864 Drosophila melanogast
gb_gss22:CNS03151	117.50	199.58	0.0376	879	AL223102 Tetraodon nigroviridis

gb_gss17:AZ163905 - 117.00 205.25 0.0182 526 ! AZ163905 SP_0074_B2_B08_T7A
gb_gss12:AO855002 + 117.00 199.29 0.0390 833 ! AO855002 CpG2052B CpIOMAGDNA
gb_est11:AI534939 + 116.50 203.06 0.0240 577 ! AI534939 SD01225.5prime SD D
gb_est14:BE253867 + 116.50 195.12 0.0666 1066 ! BE253867 601112837F1 NIH_MG

seq_name: gb_gss23:CNS04ALV

seq_documentation_block:

LOCUS CNS04ALV 898 bp DNA 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence puc-ori end of clone
095820 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL282028
VERSION AL282028.1 GI:8020362
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorphi;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A.,
Sautin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,


```
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMe 183
      ::::: ::::: ::::: ||| ||| ::::: ||:::
659 AGCAGCACACACAGCAGCAGCAGCAGTGCAGCAGCAGCAGCAGCAGCAGC 610
183 tSerAlaLeuPheSerGlyAsnThrSerSerLysGlyAlaIleLeu 200
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
609 CAGTAGC.....AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 566
200 LntHrSerAspAlaLeuThrIleThrGlyAsnGlyGlyValSerPhe 216
      ::::: ||::: ||| ||| ::::: ||::: ||::: ||::: ||:::
565 GCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 519
217 SerAspAsnThrSerSerAspSerGlyAlaIlePheThrGlyAlaSe 233
      ||| ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
518 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 469
233 rValThrIleSerAsnAlaLysValSerPheIleAspAsnLysValT 250
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
468 CACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 419
250 hrGlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCys 266
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
418 CCACCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 369
267 AlaThrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGln 283
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
368 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 319
283 tLeuLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleLeu 300
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
318 CACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 269
300 aLysLysLeuGluLeuAlaSerGlyLeuThr..... 311
      ::::: ||| ::::: ||::: ||::: ||::: ||::: ||::: ||:::
268 CCCAGGCTTCACGACAGCAGCAGCAGCTTCCTGACAGCAGCAGCAGC 219
312 LeuPheSerArgAsnSerValAsnGlyGlyThr 322
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
218 CTGCTGCCACATCATCGCGTCGACGAGCAGACT 186
```

seq_name: gb_est15:AU062062

seq_documentation_block:

LOCUS AU062062 730 bp mRNA EST 20-MAY-1999
DEFINITION AU062062 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLH423, mRNA sequence.

ACCESSION AU062062 GI:4883166

VERSION AU062062.1

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 730)

AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.

TITLE Developmental cDNA in Dictyostelium discoideum

JOURNAL Unpublished (1997)

COMMENT Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan

Email: d402h@tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers

1. 730

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SLH423"

/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

/dev_stage="slug"

BASE COUNT 247 a 137 c 150 g 196 t

alignment_scores:

Quality: 140.00 Length: 286
Ratio: 0.915 Gaps: 8
Percent Similarity: 53.497 Percent Identity: 24.126

alignment_block:

us-09-428-122-2 x AU062062 ..

Align seg 1/1 to: AU062062 from: 1 to: 730

```
78 AsnAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLe 94
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
16 AAATGCTCGCAGCAGCAGCAGCAGTGCAGCAGCAGCAGCAGCAGCAGC 65
94 uPheGlnThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSer 111
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
66 TGTATTAAACAACAATGCTGAATAAGTGAGAGTAATGCTATC..... 108
111 ervaIValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPhe 127
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
109 .....GCAGCGCTAACCACTTCAAAATCA 132
128 IleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSerCy 144
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
133 ATCGCTCAAACTGATCAGTACAGATTAAGTCAAAATGATGATGCAGC 182
144 sSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeu..... 158
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
183 ATCAGTCGAGGTTTAAACAACAACAATCATTAACAGAGAGTTCACAA 232
159 .....PheSerLysAsnPheSerThrAspAsnGlyGlyAlaIleThr 172
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
233 TTGCTAAATATCGCCAAATGCAATTCACCTGCTCTGCTGCTCAAAATCA 282
173 AlaLysThrLeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerG 189
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
283 ACAATCAAAACCAACAGTTTCATCACAACAACAACAATCA.....AGTCA 326
189 uAsnThrSerSerLysLysGlyAlaIleGlnThrSerAspAlaLeuT 206
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
327 AGGTACAAAGTGTTTATCTCTGAGCAGCTAGTATCGCTGCTGATAT 376
206 hrIleThrGlyAsnGlnGlyValSerPheSerAspAsnThrSerSer 222
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
377 CAATT...GGTAATAGTGATCAGTACCC...GGAATAACAATCAATCA 420
223 AspSerGlyAlaAlaIlePheThrGlyAlaSerValThrIleSerAsn 239
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
421 GGAAGTTTACTGTCGAATACAGTACAGTACAGTACACAGCAGCACTC 470
239 nAlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSerT 256
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
471 A.....ATTGATCCAAACAACACAA 490
256 hrThrGlyAspMetSerGlyAlaIleCysAlaThrLysThrSerThr 272
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
491 TAACTGCTTTCATCATTCAGCAATTCACAGTTCAATGATGATTA 540
273 AspThrValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsn 289
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
541 ACAACTCAACATCATTAAGTATATCATCAATCGCTGATTAATCA 590
289 nThrSerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeu 306
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
591 TCAAGCATCATCATCGCCATGCAAT..... 618
306 LaseGlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 322
      ||| ::::: ||| ::::: ||::: ||::: ||::: ||::: ||:::
618 ..... 618
323 AlaProLysGlyAlaIleAlaIleGluAspSerGlyGluLeuSerLe 339
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
```


391ThrThrGlySerSerThr 396
 850 AARTAGCTCAAGGRTCTCTTCG 872
 seq_name: gb_gss12:A0875306

seq_documentation_block:

LOCUS A0875306 520 bp DNA GSS 08-NOV-1999
 DEFINITION V122E4 min-3xHA/lacZ insertion library, strain Y2278 *Saccharomyces cerevisiae* genomic 5', DNA sequence.
 ACCESSION A0875306
 VERSION A0875306.1 GI:6287452
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM *Saccharomyces cerevisiae*
 Eukaryota; Fungi; Ascomycota; *Saccharomycetes*; *Saccharomycetales*; *Saccharomycetaceae*; *Saccharomyces*.

REFERENCE

1 (bases 1 to 520)
 Ros-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deStages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 Large-scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 Unpublished (1999)

TITLE

JOURNAL COMMENT

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 Seq primer: GCGCTTCCTTCCTTCGAAAGTAC
 Class: transposon-tagged.
 Location/Qualifiers

FEATURES

1..520
 /organism="Saccharomyces cerevisiae"
 /strain="Y2278 - S288C background, ctr(0) rho(0)"
 /db_xref="taxon:4932"
 /clone_lib="min-3xHA/lacZ Insertion Library, strain Y2278"
 /lab_host="E. coli"
 /note="Vector: pHS56-Sal. A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal, genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a min-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 140 a 102 c 169 g 109 t
 ORIGIN

alignment_scores:

Quality: 129.50 Length: 182
 Ratio: 1.126 Gaps: 6
 Percent Similarity: 63.187 Percent Identity: 27.473

alignment_block:

US-09-428-122-2 x A0875306/rev ..

Align seg 1/1 to reverse of: A0875306 from: 1 to: 520

102 ThTValAlaAlaValAlaValSerSerValValAspLysSerThr 118
 483 ACCTATCGCTCGCGCTACGACAGTTCTTA.....AAATCGACGAC 440
 118 rphelleglypheserleuserphelealeaserproglyser 135
 439 CATTAACCCGCGCAAGACCGCTACCTGCTCTCTTCTTACCTAA 390
 135 leThrThrGlyLysGlyAlaValSerSerThrThrGlySerLeu 151

389 TTTACTAGTACACAGATCATGATCTCCACTCTCATCAGCCTCA 340
 152 ThrLysAsnValSerLeuPheSerLysAspPheSerThrAspAsnG 168
 339 ACCTCTTCTGTC.....ACTACTCTCATGCTACCTCCCTCGAG 302
 168 yglAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMetSerA 185
 301 TACCGTCGCTCTAGTAGATGCTACTTCATCCATCACCACCTCATCG 252
 185 laLeuPheSerGluAsnThrSerSerLysLysLysLysAlaIleGln 201
 251 TTGCTCATGCTCATGACCACTCT..... 226
 202 SerAspAlaLeuThrThrGlyAsnGlnGlyValSerPheSerA 218
 225 TCCGACCT.....ACCTGACACTGCTGCTCTTCTCTCA 185
 218 pAsnThrSerSerAspSerGlyAlaAlaIlePheThrGlyAlaSerVal 235
 184 TCCCTGCTCAAGTCCGCTCCCTTCTCCAGCGAGTACGAGAACG 135
 235 hrLLeSerAsnAsnAlaLysValSerPheLeaSpAsnLysValThr 251
 134 CCGCTTCTTCTACAGCGCCATCTGACCTTCATCATAGGTTTGCT 85
 252 AlaSerSerThrThrGly...AspMetSerGlyGlyAlaIleCysAl 267
 84 CCTTGAGTAGTACTCGGGGTCTGACGCTCATGCA..... 49
 267 aTyThrLysThrThrAspThrLysValThrLeuThrGlyAsnGln 282
 48ACGAATAACACAGTTAAGCGCGCATTAAGAGGTAAAG 10
 seq_name: gb_est21:AW256683

seq_documentation_block:

LOCUS AW256683 727 bp mRNA EST 20-DEC-1999
 DEFINITION EST304820 KV2 Medicago truncatula cDNA clone KV2-5018, mRNA sequence.
 ACCESSION AW256683
 VERSION AW256683.1 GI:6604940
 KEYWORDS EST.
 SOURCE batrel medic.
 ORGANISM *Medicago truncatula*
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE

1 (bases 1 to 727)
 VanderBosch, K., Hult, J., Moore, J., Beremand, P., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Claven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.
 ESTs from roots of *Medicago truncatula* after *Rhizobium* inoculation Unpublished (1999)

JOURNAL COMMENT

Contact: VanderBosch K
 Department of Biology
 Texas A&M University
 College Station, TX 77843-3258, USA
 Tel: 409 845 7707
 Fax: 409 845 2891

Email: kate@mail.bio.tamu.edu
 Texas A&M EST name: T115388e
 TIGR sequence name: MTAF937K
 More information is available at: (and for clone ordering info)
 http://chrtyse.tamu.edu/medicago
 Seq primer: skmod (CTA gaa cta gta gat CC).

FEATURES

1..727
 Location/Qualifiers
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="KV2-5018"

```

/clone.lib="KV2"
/tissue.type="Seeding roots"
/dev.stage="2 days post-inoculation with Sinorhizobium
melliottii"
/lab.host="E. coli strain SOLR"
/Note: Vector: pBluescript SK-; Site1: EcoRI; Site2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unlap XR vector
from Stratagene and packaged using GigaPack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-Zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT      294 a      125 c      172 g      136 t
ORIGIN

```

```

alignment_scores:
  Quality: 127.50      length: 305
  Ratio: 0.729        Gaps: 12
  Percent Similarity: 57.377      Percent Identity: 23.279

```

```

alignment_block:
US-09-428-122-2 x AM256683/rev ..

```

```

Align seg 1/1 to reverse of: AM256683 from: 1 to: 727

```

```

97 ThrValAspAlaGlyThrValAlaGlyAlaValAlaSerSerValVa 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 ACTGTTCTGTAGGCTCACTCACTCACTCACTCACTCACTCACTCACT 678
113 IAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleAla 130
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
677 TTCTGTAGCTTGAAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 631
130 SerProGlySer...SerIleThrProGlySerGlyAlaValSerGly 145
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 CTCTAGAGTCACTCACTCACTCACTCACTCACTCACTCACTCACT 584
146 ThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSerLys 162
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
583 GTCAGATCACTTCACTCACTCACTCACTCACTCACTCACTCACTCA 534
162 nPheSerThrAspAsnGlyAlaIleThrAlaValSerLeuSerLeuT 179
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
533 AACTTCTGTAGCCACAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 496
179 hGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSerLys 195
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
495 TAGGATCACTTCACTCACTCACTCACTCACTCACTCACTCACTCA 446
196 GlyAlaIleIleGlnThrSerAspAlaLeuThrIleThr...GlyAsn 211
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
445 TCGAATTCAGTACCTCACTCACTCACTCACTCACTCACTCACTCACT 356
211 nGlyValSerPheSerAspAsnThrSerSerAspSerGlyAlaIle 228
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
395 TACAGTGTACAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 346
228 IePheThrGluAlaSerValThrIleSerAsnAsnAlaValSerPhe 244
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
345 TAGCCACAGANTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 317
245 IleAspSerValThrGlyAlaSerSerSerThrThrGlyAspMetSe 261
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
316 .....GTAGCTCAAGCTTCACTCACTCACTCACTCACTCACTCACT 287
261 rGlyGlyAlaIleGlyAlaValThrLysThrSerThrAspThrLysVal 278
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
286 .....TCTTTTACGTTTCTGTAGCTTCACTCACTCACTCACTCACT 253
278 eutThrGlyAsnGlnMetLeuPheSerAsnAsnThrSerThrThrAla 294
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
252 CCACCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA 203

```

```

295 GlyValAlaIleThrValLysLysLeuGluAlaSerGlyGlyLeuThr 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GGTTCCTCTTTACCGCTT.....TCTGTACCTCACTCACTCACTGC 162
311 rIeUPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGlyGly 328
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
161 CACAGGTTCTTCTTTTACCGCTTCTGTAGCTCACT.....T 124
328 lAlleAlaIleGluAspSerGlyGluLeuSerLeuSerAlaAspSerGly 344
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
123 CACTACCAAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA 77
345 AspIleValPheLeuGlyAsnThrValThrSerThrThrProGlyThr 361
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
76 .....GTAGCTACAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 53
361 nArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAlaLeuArg 378
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
52 .....GTTCTGTAGATCACTCA..... 32
378 eAlaAlaGlyArgAlaIleThrPheThrAspProLeuThrThrGlySer 394
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
31 .....GTACACACAGTTTCG 17
395 SerThrThrValThr 399
16 TCTTTACGTTTCT 2
seq_name: gb_gss9:AO645402
seq documentation block: 672 bp DNA GSS 08-JUL-1999
LOCUS AO645402
DEFINITION RC193-DpniI-29D12.1J RC193-DpniI Trypanosoma brucei genomic clone
ACCESSION RC193-DpniI-29D12, DNA sequence.
VERSION AO645402.1 GI:5122112
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 673)
El-Sayed,N., Zhao,S., Zhao,H., Gull,S., Suh,E., Malek,J., Fujii,C.,
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPI-93
Library for gene discovery and sequence-ready map construction
Unpublished (1999)
Other GSSs: RPI93-DpniI-29D12.TV
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
7712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@igrr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mdb/tbdb/.
seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..673
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:3691"
/clone="RPI93-DpniI-29D12"
/Note="Vector: pBAC3.6; Site_1: Bam HI; Site_2: Bam HI;
Constructed for The Institute for Genomic Research by
Bohai Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma

```


[illegible]

seq_name:	gb_est18:AV403135
seq_documentation_block:	500 bp
LOCUS	AV403135 mRNA EST
DEFINITION	AV403135 Bombyx mori middle silk gland 5th instar larva Bombyx mori
CDNA clone	msg90085 t3, mRNA sequence.
ACCESSION	AV403135.1
VERSION	GI:6907223
FEATURES	EST

```

ORGANISM      Bombyx mori
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
               Bombycoidea; Bombycidae; Bombyx.
REFERENCE     Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
AUTHORS       Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE         Bombyx mori cDNA
JOURNAL       Unpublished (2000)
COMMENT       Contact: Mita K
               Genome Research Group
               National Institute of Radiological Sciences
               Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
               Email: kmita@nirs.go.jp
               method:uni-directional, sequence direction:sequenced from T3 primer
               (5' -> 3')
               Project=Silkworm Genome Program in MAFF, and Research for the
               future program in USFS, see /silkbases/,
               Future program in USFS, see /silkbases/, for whole ESTdb.
               http://www.ab.a.u-tokyo.ac.jp/silkbases/,
               Location/Qualifiers
                   length=500
                   organism="Bombyx mori"
                   db_xref=taxon:7091
                   clone_lib=Bombyx mori middle silkgland 5th-instar larva
                   sex=female/male mixed
                   tissue-type=middle silkgland
                   dev_stage=5th-instar larva
                   note='donated by Dr. Sehmal, Czech'
BASE COUNT   142 a 129 c 125 g 103 t 1 others
ORIGIN
Alignment-scores:
    Quality: 126.00 Length: 192
    Ratio: 1.156 Gaps:
    Percent Similarity: 56.771 Percent Identity: 28.125
Alignment-block:
US-09-428-122-2 x AV403135 ..
Align seq 1/1 to: AV403135 from: 1 to: 500
..109 AsrSeSerValValAspySerThrThPheIleGlyPheSsrLe 125
:::|||||:::|||||:::|||||:::|||||:::|||||
16 AGTCACAGTAACTCATGTCAACACACAGACTTGACAGATCCATC 65
125 userPhellealeaserProglyserGlyserIlethrTrcAlgyALAY 142
||| ||| ||| ::::: ||| |||:::
66 CGGGTAGCACACCTTAATGATACAGTTGCCACAGTCGTATAGGAAGT 115

```

Fri NOV 24 13:50:17 2000

us-09-428-122-2.rst

Page 8

```

142  aIsErCySeSrThrGlyLeuLeuSerLeuThrIysAsnValSerLeuLeu 158
      ||||| ||||||||| || |||
146  TATCA...TCCACGGCGACGTCCTACATACT..... 144
159  PheSerIysAsnPhSeSerThrAspAsnGlyAlaIleThrAlaIysTh 175
      :||| :|||:||||| :|||:||||| :|||:|||||
145  .....GATCCAGACACACAGCTTCGAGCATCCAGATCCAGATCGGCGG 185
175  rIeuSerLeuThrGlyThrThMetSerAlaIeuPhSeSerGlnAspThS 192
      ||| ||| :||| :|||:||||| :|||:|||||
186  TACACACACTTATGATAGATGATCCAAAGT.....CGTATGAAGAG 229
192  eErSerIysGlyGlyAlaIleGlnThrSerAspLeuThrIleThr 208
      :|||:||||| :|||:||||| :|||:|||||
230  TATCATCCACGGCGAGTCCAGTATACATGATSCAAGCAGCAGACTTAA 279
209  GlysAngin.....GlyGlyValSerPhSeSerAspAsnThrSerSe 222
      :|||:||||| :|||:||||| :|||:|||||
280  GGATCCAGTACATCCGGCGGTAGCAGCATATGATATACTATACAAACA 329
222  rAspSerGlyAlaAlaIlePhThrGluAlaSerValThrIleSerAsn 239
      :|||:||||| :|||:||||| :|||:|||||
330  TCCTATGAGATGTATGCGCCACGCGCACT.....TCCAGTA 367
239  snAlaIysValSerPheIleAspAsnIysValThrGlyAlaSerSer 255
      :|||:||||| :|||:||||| :|||:|||||
368  ACACGTATCCAGACACCAAGAGATGCCACGCGCGGTAGAGAC 417
256  ThrThrGlyAspMetSerGlyValAlaIleCysAlaTyrIlyThrSerTh 272
      ||| ||| :||| :|||:||||| :|||:|||||
418  ACAGAGATATATGCT.....ACAGCCA 443
272  rAspThrIysValThrLeuThrGlyAsnGlnMetLeuLeuPhSeSerAsn 289
      ||| :|||:||||| :|||:|||||
444  TGTGTAAAGCGTAA.....TCCACCG 466
289  snThrSerThrThrAlaGlyGlyAla 297
      :|||:||||| :|||:||||| :|||:|||||
467  ACGGTTCACGACACAGTGGAGACT 492

```

Locus AV403126 650 bp mRNA EST 06-FEB-2000
Definition Bombyx mori middle silk gland 5th-instar larva Bombyx mori
Accession AF403126 CDNA clone msgv072 T3, mRNA sequence.
Version AV403126.1 GI:6907214
Keywords EST.
Source domestic silkworm.
Organism Bombyx mori
Reference Pempkita, Melazoa; Athropoda; Tracheata; Hexapoda; Insecta;
 Lepidoptera; Endopterygota; Lepidoptera; Glossata; Dityrslatidae; Bombyloidea; Bombycidae; Bombyx.
AUTHOR Orlimyo M., Shimada,T., Okano,K. and Maeda,S.
Title CDNA (2000)
Journal K Group
 Institute of Radiological Sciences
 Chiba University
 Chiba 263-8555, Japan
 E-mail: sequnce.direction@sequenced from T3 primers
 (see "Silkbases")
 ac.jp/silkbases/>, for whole ESTdb.

BASE COUNT	ORIGIN
187 a	175 c 164 g 124 t

```
alignment_scores:
  Quality: 126.00
  Ratio: 1.156
  Percent Similarity: 56.771
  Length: 192
  Gaps: 7
  Percent Identity: 28.125
```

```
alignment_block:
    ne = 00-439-133-3
```

Align seg 1/1 to: AV403126 from: 1 to: 650

```

109 AsnSerSerValValAspLysSerThrThrPheLeuGlyPheSerSer 125
110 ||||| ||||| ||||| ||||| |||||
111 AGTCCAGTACCATCATCAACAGACAGACCTTCGAGATCCAGTCAATC 65
125 userPheLeuLaserProGlySerSerLeuThrThrLeuThrGlyAlaVal 142
126 ||||| ||||| ||||| ||||| |||||
127 CGCGCGTAGCCAGCATTAGGTACACTTCCACAGTCCGATGAGGAGG 115
142 alSerCysSerThrGlySerLeuSerLeuThrThrGlyAsnValSerLeu 158
143 ||||| ||||| ||||| ||||| |||||
144 TATCA...TCCACCGGCACCTCCAGTACACT... 144
159 PheSerLysAsnPheSerThrAspAsnGlyAlaAlaLeuThrAlaLysTh 175
160 ||||| ||||| ||||| ||||| |||||
161 ..GATCCAGCCAGCAGACCTTCGAGATCCAGTCAATCAGCGCGCG 185
175 LeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerGlyAsnThrS 192
176 ||||| ||||| ||||| ||||| |||||
177 TAGCAGCAGCATTATGATATACAGTCCAAAGT...CGATGAGGAGTG 229
192 eSerLysLysGlyAlaAlaLeuGlnThrSerAspAlaLeuThrLeuThr 208
193 ||||| ||||| ||||| ||||| |||||
194 TATCATCCACGCGCAGCTTCAGTACACAGTCAAGACAGACGACTTACA 279
209 GlyAsnGln.....GlyGluValSerPheSerAspAsnThrSerSe 222
210 ||||| ||||| ||||| ||||| |||||
211 GGATCCACTACATCCCGCGTAGACAGCACTTATGATATAGCTCAACGA 329
222 rAspSerGlyAlaAlaLeuPheThrGluLaserValThrLaserAsna 239
223 ||||| ||||| ||||| ||||| |||||
224 TCGATGATGAGAGCTATTGGCCACTGGCAGT.....TCCACTA 367
239 snAlaLysValSerPheLeuAspAsnLysValThrGlyAlaSerSerSer 255
240 ||||| ||||| ||||| ||||| |||||
241 ACACGATGACACAGCACACAGAAATACACACACCTCCGCTGTATGACG 417
256 ThrThrGlyAspMetSerGlyAlaAlaLeuCysAlaThrLysThrSerTh 272
257 ||||| ||||| ||||| ||||| |||||
258 ACTGAGGATATATAGTTC.....AGTAGCCA 443
272 rAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsna 289
273 ||||| ||||| ||||| ||||| |||||
274 TGTATGAGAGGATACA.....TCCACCG 466
289 snThrSerThrThrAlaGlyAla 297
290 ||||| ||||| ||||| ||||| |||||
291 ACGCTTCACGACCAATGATGAGAGACT 492
seq_name: gb_est9:AI16701
seq_documentation_block:
LOCUS AI16701 693 bp mRNA EST 03-DEC-1998
DEFINITION xylem:est.50 Poplar xylem Lambda ZAPII library
subsp. trichocarpa cDNA 5', mRNA sequence Populus balsamifera

```

```
/clone_11b="GH Drosophila melanogaster head POT2"  
/sex="male and female"
```

```

/rev stage="adult"
/ab host="DB5" - "alpha"
/Note="Organ: head; Vector: pOT2; Site:1; Ecot: Site:2;
XhoI: sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT      133 a      207 c      129 g      144 t
ORIGIN

```

```

Alignment scores:
Quality: 124.00      Length: 263
Ratio: 0.879         Gaps: 13
Percent Similarity: 53.612      Percent Identity: 25.095

```

alignment_block:
US-09-428-122-2 x AIS17569/rev ..

Align seg 1/1 to reverse of: AIS17569 from: 1 to: 613

```

68  ThrGlyThrAlaIleThrlYserCysPheAsnThrLysGlyAspLe 84
|||||.....|
607  ACTGGCGCTTCATGCGCAGCTTCT.....GATAG 579
84  urThrPheThrlGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAlaG 101
|||  |||.....|
578  TACAGAACATCGATGAGTTC.....GAG 553
101  lYThrValAlaGlyAlaAlaValAsnSerSerValAlaAspLysSerThr 117
|||||.....|
552  GCACCTGCCGACGAGTGTCC.....GTGGTCTCGGAAACGAA 515
118  ThrPheIleGlyPheSerLeuSerPheIleAlaSerProGlySerLe 134
|||  |||.....|
514  GCCGATGTTGGAGTAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 465
134  rIleThrlThrlGlyLysGly...AlaValSerCysSerThrlGlySerLeuS 150
|||||.....|
464  AGCAGGTGCATCAGAGGCGCGCATACAGGAGCACTTGTGATAT 415
150  erLeuThrlLysAsnValSerLeuLeuPheSerLysAsnPhSerThrAsp 166
|||  |||.....|
414  CTAAAGCTGCTATGCTACT.....GCTCTGAT 386
167  AsnGlyGlyAlaIleThrlAlaLysThrLeuSerLeuThrlGlyThrIle 183
|||||.....|
385  GCTGGTGGCTGCTCTCAGAGTTCCTTCAAGCTACAGCTGAGAGG 336
183  tSerAlaLeuPheSerCysLysnThrSerSerLysLysGlyGlyAlaIleG 200
|||||.....|
335  TTGGGAGGCC.....ACAGGAGGAGAGACAGGCTGGCTAGT 238
200  lnrThrSer.....LeuThrlIleThrlGly...AsnGlnGlyValSer 215
|||  |||.....|
297  CG.....AACGACCTAGAGAGGCGGTATGAGCAGTC... 251
|||  |||.....|
serAspSerGlyAlaAla...IlePheThrG 231
|||||.....|
TAAATTCGCGAAGACAGCAGTGCAGTGCATC 213
AsnAlaLysValSerPheIleAspAsnL 248
|||||.....|
TGCAGGAGACTGCTGTCGTCGACAGCA 166
ThrlGlyAspMetSerGlyAla 264
|||||.....|
AGGAGGTACAGCGGTGTGT 116
LysValThrlLeuThrlGlyAs 281
|||||.....|
92
ThrlAlaGlyGlyAlaI 298
|||||.....|

```

```

91  .....ACTTCGGGCTTCTCCACACAGCTGTCGCA 58
298  leryThrlLysLeuGlnLeuAlaSerGlyLeuThrlLeuPheSer 314
57  T.....|
315  ArgAsnSerValAsnGlyThrlAlaProLysGlyGly 327
|||  |||.....|
55  .....GTTCTGGCTCTGGGCTGAGAGTACGAGT 26
seq_name: gb_gss24:CNS05GDN

```

```

seq_documentation_block:
CNS05GDN      1068 bp      DNA      GSS      26-MAY-2000
LOCUS      Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION      014C07 of library B from Tetradon nigroviridis, genomic survey
sequence.

```

```

ACCESSION      AL336164
VERSION      AL336164.1 GI:8229922
KEYWORDS      GSS: genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM      Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.

```

```

REFERENCE      1 (bases 1 to 1068)
AUTHORS      Roest-Crolius,R., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

```

```

TITLE      Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL      2 (bases 1 to 1068)
UNPUBLISHED

```

```

REFERENCE      Roest-Crolius,R., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
HUMAN GENOME NUMBER ESTIMATE PROVIDED BY GENOME WIDE ANALYSIS USING
Tetradon nigroviridis DNA sequence
UNPUBLISHED
3 (bases 1 to 1068)

```

```

FEATURES
source
1..1068
location/Qualifiers
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="B"
/clone_id="B"
BASE COUNT      234 a      236 c      233 g      347 t      18 others
ORIGIN

```

```

Alignment scores:
Quality: 123.50      Length: 322
Ratio: 0.627         Gaps: 10
Percent Similarity: 61.180      Percent Identity: 21.739

```

alignment_block:
US-09-428-122-2 x CNS05GDN ..

Align seg 1/1 to: CNS05GDN from: 1 to: 1068

```

26  ValLeuAspSerAlaSerPheAspGlyAsnLysAsnGlyAsnPhese 42
|||  |||.....|
90  GTGGCGAAGCTTCTCTGCTCAGAGAGCTAAGTGTCTCAGAGCT 139
|||||.....|

```

Fri Nov 24 13:50:17 2000

```

338 rleuserAlaaspSer 343
1 .....
905 TTCTGTCTCTCTCC 920

seq_name: gb_gss9:A0640209
seq_documentation_block:
LOCUS      A0640209      579 bp      DNA      GSS      08-JUL-1999
DEFINITION 927Pl-2F1-TV 927Pl Trypanosoma brucei genomic clone 927Pl-2F1, DNA
sequence.
ACCESSION  A0640209
VERSION    A0640209.1 GI:5116919
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Trypanosoma brucei.
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE  1 (bases 1 to 579)
            El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
            Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donalson, J.,
            Fraser, C. and Adams, M.
            Determination of clone end sequences from Trypanosoma brucei TREU
            927/4 Pl library
            Unpublished (1999)
            Other GSSs: 927Pl-2F1.TP
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            For clone/filer availability, please contact Sara Melville
            (sm560@ole.bio.cam.ac.uk). Pl end sequences search page:
            http://www.tigr.org/cdb/mbd/ibdb/.
            Seq primer: T7
            Class: Pl ends.
            Location/Qualifiers
                1..578
                /organism="Trypanosoma brucei"
                /strain="TREU927/4"
                /db_xref="taxon:5691"
                /clone="927Pl-2F1"
                /clone_1b="927Pl"
                /note="Vector: pAD10sacBII; Site 1: Bam HI. Constructed by
                Sara Melville, University of Cambridge, UK and Nancy
                Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was
                isolated from Trypanosoma brucei (stock TREU927/4) and
                partially digested with sau 3A1. DNA fragments were cloned
                into the Bam HI site of pAD10sacBII vector (Genbank
                accession U09128). The average insert size is 65 kb.
                Coverage: approx 4.4 x the haploid non-mitochondosomal
                genome"

```

```

BASE COUNT      114 a      101 g      183 t
ORIGIN
alignment_scores:
    Quality: 123.00      Length: 170
    Ratio: 1.025      Gaps: 4
    Percent Similarity: 70.588      Percent Identity: 27.059

```

```

alignment_block:
US-09-428-122-2 x A0640209
Align seg 1/1 to: A0640209 from: 1 to: 579

```

```

95 pNegIntrValaAspAlaGlyThrValaIaGlyAlaValaIaAspSer 111
|||||
77 TTCAATAGAGTAGTAGACATCTCTGCTCAAGTGAAGCTATATC 126
|||||
111 rValValaAspLysSerThrThrPheIleGlyPheSer 124
|||||

```


519 ..CAATCTCCAGTAGTTCATCTCAGGCTCTACTACAC
 275 ..SerGlyValAlaIleCysAlaTyrIlePheSerThrAspThr 275
 564 ..ATCTCCAGGTAGTTCATCTCAGGCTCTACTACAC 601
 292 ThrThrAlaGlyGlyAlaIleTyrValLysLysLeuGluLeuAlaSerG1 308
 602 TCCTCTCTGTTGGTGGCGGCGGCTCTCCACCAGGTACAGTCAATCTCC 651
 308 yGlyLeuThrLeuPheSerArg...AsnSerValAsnGlyLys..... 321
 652 AGGTACTTCATCTCAGGCTCTCAGGCTCTCTGTTGGTGGCGGCT 701
 322ThrAlaProLysGlyGlyAlaIleAlaIleGluAsp 333
 702 CTTCACCGCTGCGGTAGCGGTGAGACAGGTTCACAGTGTAT 742
 seq_name: gb_gss23:CNS03TF7
 seq_documentation_block: 918 bp DNA 17-MAY-2000
 LOCUS CNS03TF7
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-Orl end of clone
 056109 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL259756.1 GI:7980768
 VERSION 1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Euplaterygii; Ctenosquamata; Acanthomorphia; Eucanthomorphia;
 Holacanthopterygii; Acanthopterygii; Perciformes;
 Tetraodontiformes; Tetraodontidae; Tetraodon.
 1 (bases 1 to 918)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
 Bouneau,B., Billault,A., Queller,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 2 (bases 1 to 918)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,B., Fisher,C.,
 Bernot,A., Fizes,C., Mincker,P., Brottier,F., Queller,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 918)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/Genbank/DDBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetradon.
 FEATURES
 source
 1. 918
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="056109"
 /clone_lib="G"
 /note="Genoscope sequence ID: COB9056CF05SP1-end :
 PUC-Orl"
 BASE COUNT 293 a 324 c 136 g 159 t 6 others
 ORIGIN

alignment_scores: 121.00 Length: 301
 Ratio: 0.752 Gaps: 12
 Percent Similarity: 53.488 Percent Identity: 23.586
 alignment_block:
 US-09-428-122-2 x CNS03TF7 ..
 Align seq 1/1 to: CNS03TF7 from: 1 to: 918
 100 AlAGlyThrValAlaGlyAlaValAsnSerSerValAlaLysPylse 116
 72 GCAGCTCTGTACACCATCAGCAGCAACACTGAGTATCAACACTAC 121
 116 rThrThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlyS 133
 122 AACTACA.....GAACTGGAG 138
 133 eSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeu 149
 139 GCACCAACAACCTTGCTGCCAACAAGACAGCAACCAACACTGTTGA 188
 189 SerLeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrS 166
 166 PASNGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThm 183
 236 GACTTCATCAGCAGAACTTATCCACCATCAACCCACCAACCAAAACCA 285
 183 eSerIleLeuPheSerGluAsnThrSerSerLysLysGlyAlaIle 199
 286 CAGCAAGACGCTCAACCAACCAACATCA.....GGTACCTC 326
 200 GlnThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerP 216
 327 ACTACATAGAA.....ACGCAACAACCTTCAG.....ACCTT 361
 216 eSerAspAsnThrSerSerAspSerGlyAlaIlePheThrGluAs 233
 362 CTACCATCTATCATCAGATCAAGT..... 389
 233 eValThrIleSerAsnAsnAlaLysValSerPheLeaSpAsnLysVal 249
 390 ..GTACAACTAGTAACACAMCTCCACTCCACTCTGCTCCACCTG 437
 250 ThrGlyAlaSerSer.....ThrTh 257
 438 ACTCCAAACCAAGTGCAGCTCTAAACACCAACCAACATGAGAGTTAC 487
 257 rGlyAspMetSerGlyAlaIleCysAlaTyrIlePheSerThrAspT 274
 488 AAMCCCAATCATCAGTAGTACCTCACTCAATGGAAGGCAAACTT 537
 274 hrLysValThrLeuThrGlyAsn.....GlnMetLeuLeuPheSer 287
 538 CTGAGCGCTTTTAACATCCGACTCCACAACAGAGGTACAACTAGT 587
 288 AsnAsnThrSerThr.....ThrAlaGlyGlyAl 297
 588 AAGGAGCTCTCACTCACTCTACTCTCACTCACTCACTCACTCACTG 637
 297 AleIleThrLysLysLeuGluLeuAlaSerGlyLysLeuThrLeuPheS 314
 638 AGCTCTGTACCA...TCTTACGACAACTGAACTTACCTCCG 684
 314 eArgAsnSerValAsnGlyLysThrAlaProLysGlyAlaIleAla 330
 685 CACCTCCAGCAAGCAAGGAGGAGCAAGAGTACGCGTTCTCTGCA 734
 331 IleGluAspSerGlyLysLeuSerLeuSerAlaAspSerGlyAspIle 347
 735 GTAGCAGCAACCAAGCAACCAACCAACTTCTGTTTATCAGACACCAAT 784

347 lPheLeuGlyAsnThrValIhrSerThrThrProGlyThrAsnArgSer. 363
: : : : :
785 C.....ACGAGGACTGTGTAGAAAGCAGACACTGGAGRACCACTGCAGCCG 828
364SerIleAspLeuGlyThrSerAlaLysMethr 374
829 CCAGCTCAGACACTCCAGACAGGGGTATATACTTCAGCCTCCAGACCC 878
375 Ala 375
111
879 GCT 881

470 ACGGTACAGAG
258 YASW
523
Fri Nov 24 13:50:17 2000